PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/53, 33/566

A1

(11) International Publication Number: WO 99/40435

(43) International Publication Date: 12 August 1999 (12.08.99)

(21) International Application Number: PCT/US99/02612

(22) International Filing Date: 6 February 1999 (06.02.99)

(30) Priority Data:
60/074,070 9 February 1998 (09.02.98) US

(71)(72) Applicant and Inventor: NETZER, William, J. [US/US]; 1967 Ocean Avenue, 4D, Brooklyn, NY 11230 (US).

(74) Agents: WARBURG, Richard, J. et al.; Lyon & Lyon LLP, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071-2066 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: PROTEIN FOLDING INHIBITORS

(57) Abstract

The subject disclosure relates to strategies and methods for the discovery, development, and use of drugs and drug lead molecules that inhibit protein folding (folding inhibitors). These can consist of small organic molecules that bind nascent polypeptides selectively within cells during their synthesis on ribosomes and/or before folding of the protein is completed, and by virtue of this activity inhibit the target polypeptide from folding to its native state which is otherwise responsible for its biological activities. Methods for discovering folding inhibitors include, for example: 1) the design of specific peptide and non-peptide inhibitors, 2) the identification of suitable chemistries for the synthesis of combinatorial libraries of small organic molecules, 3) aptamers, and 4) effective screening methods. The present invention also relates to methods of enhancing the potencies of said compounds, which are expected to have extraordinary medicinal properties.

AQ3

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DESCRIPTION

PROTEIN FOLDING INHIBITORS

BACKGROUND

This invention relates to the field of the identification and use of therapeutic compounds, in particular to the identification and use of compounds active on particular proteins.

The following description is provided to assist the understanding of the reader. None of the information provided or the references cited are admitted to be prior art to the present invention.

Proteins consist of linear sequences of amino acids that under physiological conditions fold into highly ordered three dimensional structures referred to as the folded or native state. Folding is dictated by the amino acid sequence of the polypeptide (primary structure). In general the folded, or native state of a protein molecule imparts the protein's biological activity, which may, for example, be the specific catalytic activity of an enzyme, the ligand binding and signaling activity of a receptor or signal transduction molecule, the function of a transport molecule, or the mechano-chemical properties of a structural or cytoskeletal protein. Thus, inhibition of folding results in loss of biological activity.

In general, proteins consist of folded units called domains. These are defined as independently foldable or semi-independently foldable polypeptide chains or chain segments of characteristic folds (tertiary structure) formed by the positioning of secondary structural elements (e.g., alpha helices, beta strands and sheets, loops) in space relative to one another. Proteins contain one or more domains. The folding of a multidomain protein also involves the orientation of different domains relative to one another, where such orientations are fixed.

The folding of a polypeptide into a compact highly ordered domain is a cooperative process governed by interactions between numerous parts of the polypeptide chain and by interactions between the polypeptide and its solvent. In general, domains consist of an outer shell containing predominantly hydrophilic amino acids, which surrounds a core of predominantly hydrophobic amino acids.

Domains which form interfaces with other domains generally contain hydrophobic amino acids at the domain-domain interface. Mutational studies have demonstrated that the native fold of a domain can be perturbed by changes in packing volume and/or by changes in charge distribution within the core region. Specific domains, or folds, are generally found to occur in more than one protein even where amino acid sequences differ. Very often specific functions are localized to a given domain in multidomain proteins. Additionally, active sites of many enzymes occur in clefts situated at the interface of two or more domains of multidomain proteins.

Proteins are synthesized within cells on ribosomes and, in general, must fold to assume biological activity. Folding within cells generally requires the assistance of molecular chaperone proteins (that are possessed by all cells). However, the main job of chaperones is to prevent the aggregation of unfolded protein. Chaperones of the Hsp90 class also function to regulate the activity of numerous signal transduction proteins by binding to relatively unstable domains within these proteins. It is still the amino acid sequence of the protein that determines its folded structure.

In eukaryotic cells, the folding of a polypeptide chain requires that a complete protein domain be synthesized and sufficiently extruded from the ribosome so that topological restriction of folding by the ribosome does not occur. As a result, single domain proteins fold post-translationally, i.e., after the polypeptide chain is completed and released by the ribosome. While ribosome bound, such proteins remain unfolded. Multiple domain proteins fold in a sequential manner with N-terminal domains folding independently, and ahead of C-terminal domains that are not completely synthesized. Thus, folding is sequential and co-translational. This is characteristic of multi-domain cytosolic proteins and such proteins that are transported to the endoplasmic reticulum. In a multi-domain protein, a single N-terminal protein domain remains unfolded on the ribosome until completion of its synthesis and the synthesis of at least part of an adjacent C-terminal domain. The C-terminal domain of such a protein folds after it is synthesized and released from the ribosome.

Proteins or protein domains that interact with chaperones are in an unfolded state and can only fold when released from the chaperone(s). This is because, in general, chaperones bind unfolded proteins.

In bacteria, cytosolic proteins appear to fold post-translationally, i.e., folding to the native state may not occur, for many proteins, until the polypeptide chain is completed and released from the ribosome. In this case, a protein will also remain unfolded as long as it is bound to the ribosome. Additionally, proteins bound to bacterial chaperones are in an unfolded state.

In all cells, the fate of a protein unable to fold is rapid degradation or formation of an insoluble aggregate. In eukaryotes, degradation is accomplished mainly by the ubiquitin system. In bacteria, other proteolytic systems assume this function. Chaperonins of the Hsc 60 class may sometimes refold misfolded proteins, but if a protein is unable to fold, it will be transferred to degradation pathways.

During times of stress, such as heat shock, proteins have a tendency to unfold. In eukaryotes, such proteins may be refolded principally by chaperones of the Hsp 90 and Hsp 70 classes. Binding of proteins to these chaperones occurs after stress-induced unfolding. (Proteins may also bind to these chaperone systems during de novo folding.) Proteins unable to refold (or fold in the context of synthesis) are degraded.

Classification of Drug Molecules

To a great extent, medicinal chemistry concerns the binding of a ligand to a biological receptor, generally a protein, where binding is characterized by specificity and affinity, and is based on structural aspects of the ligand (known as the pharmacophore) which recognize (or are recognized by) complementary aspects of the unique three dimensional structure, or native fold of the protein host or target (binding site). Most drugs are ligands which contain pharmacophores that, to varying extents, mimic natural ligands of specific proteins such as enzymes and receptors. As such, the drug is able to bind to a target protein's ligand binding or active site. In some cases binding to a natural ligand binding site may be accomplished by a drug molecule that does not resemble a natural ligand. Also, some drug-like molecules may bind proteins at sites other than ligand binding sites and may modulate the activity of the protein through this binding.

Drug ligands that bind proteins may inhibit the protein's activity (antagonists) by preventing access to a binding site directly or indirectly by a natural

or alternative ligand, or they may stimulate activity (agonists) by stimulating a signal similar to that induced by a natural, agonist ligand. However, all of these drugs require that their protein targets be folded so that unique tertiary structure is available to mediate the recognition process, and, in the case of agonists, to mediate transmission of a biochemical signal. More recently, antisense technologies have targeted (inhibited) the flow of genetic information from nucleic acid to protein through agents that bind to RNA and prevent its translation into protein. No drug has previously been conceived to inhibit the process of protein folding. The novel drugs and drug lead compounds whose discovery are the goal of this invention are designed however to fulfill this role, and as will be made clear will provide unique benefits and major improvements over extant drugs. Additionally, their means of derivation will yield major improvements over current drug discovery technologies.

Drug Development and Lead Compound Discovery

Much of modern drug discovery/development involves the identification of drug leads. These are compounds which are predicted or demonstrated to have biological or medicinal properties of interest. Occasionally a lead compound discovered in a pharmacological assay or screen is utilized as a clinically prescribed drug without undergoing significant chemical modification (e.g., lovastatin). However, most lead compounds undergo substantial chemical modification after their identification which results in improved drug-like properties. An example is the use of noradrenaline and ephedrine as lead compounds for alpha-1 adrenergic agonists, such as salbutamol. In general, a lead compound can be more accurately described as one which has a property or properties of medicinal interest but whose properties do not necessarily match all of the properties of an anticipated or sought after drug. Lead compounds generally contain a chemical core structure, or pharmacophore that is responsible for a key pharmacological property. Though often chemically modified during pharmaceutical development, the pharmacophore provides a basis for further development.

The ability to bind a receptor and elicit a particular response is only one aspect of a drug. The complete behavior of a drug in the body (i.e., pharmacodynamics) includes its ability to reach its target (i.e., bioavailability) which depends, in part, on how it is distributed throughout the body and for how long (i.e.,

pharmacokinetics). Other important features include drug metabolism, toxicity and side effects.

SUMMARY OF THE INVENTION

The present invention relates to the discovery and use of molecules, including small organic molecules, that inhibit the folding of target proteins and to methods for the screening, discovery and optimization of such molecules. The molecules have the property of specifically recognizing and binding to target unfolded or partially folded proteins. This inhibits folding of the target protein to its native, and biologically active state, and also results in the target protein's rapid degradation or aggregation. Preferably for clinical use of folding inhibitors, the molecules readily enter cells because their primary site of action will be the translating ribosome or other cellular compartments where proteins have folded post-translationally or become unfolded (e.g., protein bound to chaperones). However, it is not necessary for initial lead compounds which are shown to inhibit protein folding in an assay to have the property of intracellular delivery (though it is advantageous).

Folding inhibitors can be regarded as falling into two classes: 1) Inhibitors that bind short peptides within unfolded proteins or within unfolded protein domains ("class 1"); and 2) inhibitors that bind to pockets or other folded structural features of a protein domain (or in other instances to a subdomain consisting of a cluster of secondary structures) at the interface of that domain (or subdomain) and one or more other domains (or subdomains) of the protein.. Such class 2 inhibitors take advantage of the fact that domains of a multidomain protein fold independently (the subdomains, or secondary structural units within individual domains are also known to form before folding to native tertiary structure occurs).

One group of class 1 inhibitors includes small organic molecules. These can, for example, be derived by screening a combinatorial library of such compounds or collected library of compounds, utilizing (as probes for screening) either short peptides chosen from the amino acid sequence of a target protein or an essentially complete. and denatured (unfolded) polypeptide chain of the target protein. In some embodiments, such a compound or plurality of compounds that bind(s) to at least

two different sites within a domain of the target polypeptide is utilized. In the case where two or more compounds are used, the compounds can be chemically ligated to form a bivalent or multivalent polypeptide binding compound.

Class 1 inhibitors can, for example, include various formulations of short peptides that bind sequences within the larger unfolded target protein, or other peptides that have been derived, for example from a combinatorial library. Additionally, class 1 inhibitors may include aptamers or other types of compounds that bind target sequences within an unfolded protein.

Class 2 inhibitors can include small organic molecules derived through computer-aided computational methods, which generally involve either the scanning of an existing or theoretical (i.e., virtual) chemical library or which are derived by design and computation involving the surface of a target protein. Such molecules will bind to an interfacial surface defined by adjacent folded protein domains (or subdomains), and will also, in some instances, include peptidomimetic compounds.

The inhibition of folding to stable tertiary structure by either class of inhibitor will result in a) loss of activity (antagonism) of the inhibited protein and/or b) rapid degradation of that protein within cells.

For the development of Class 1 inhibitors, it is highly beneficial that the amino acid sequence (or part thereof) of the target protein be known. Although useful, detailed structural information is not obligatory. This immediately extends the scope of available pharmacological targets and allows one to take advantage of the enormous and growing protein sequence and genomic databases. Furthermore, peptide segments of a target protein are sufficient for the discovery or design of class 1 folding inhibitors and therefore the target protein itself need not be present at this stage.

Additionally, both classes of inhibitors can, and generally will, be unrelated to any of the natural ligands of the target protein. This means that other receptors that interact with natural and artificial ligands of the target protein will not be acted upon by the folding inhibitor provided that the proteins do not posses the specific amino acid sequences addressed by the folding inhibitor (or similar interfacial surfaces for class 2 inhibitors). This allows discrimination of targets based on subtle differences in amino acid composition. There are also numerous potential target sites within any target protein, and inhibitors that address these different peptide

sequences will have different pharmacophores. The ability to target multiple sites is especially useful in treating infectious diseases and cancer, where mutations in either pathogens or tumor cells can result in resistance to single drugs. By addressing multiple targets, the probability of "escape" by mutation is equal to or less than the product of the individual frequencies of each mutation; "less than" results because mutations in a protein also increase the likelihood that the protein will be non-functional (therefore certain "escape" mutations will not be allowed).

Organic molecules, preferably small molecules, that bind to proteins after synthesis yet before they fold into their native, or active, configuration, are selected by the method of the present invention. Such molecules are included in the term "protein folding inhibitors".

As described herein, a variety of techniques can be used to identify protein folding inhibitors, including methods involving synthesis of a target protein in the presence of test compound, binding of test compound to a peptide or other site of a protein, and computational methods for identifying binding compounds.

Thus, in a first aspect, the invention provides a method for identifying a protein folding inhibitor, involving contacting a protein biosynthetic system under protein synthesis conditions with at least one test compound, and determining whether the test compound increases the ratio of unfolded protein to folded protein. An increase in that ratio is indicative that the test compound is a protein folding inhibitor. As described in the Detailed Description below, it is generally helpful to include the use of controls to confirm that observed effects are actually due to inhibition of protein folding. In preferred embodiments of this and following identification methods, experimental controls are included. The method can be used to test or confirm the effect of a single particular test compound (e.g., one which is believed to be a protein folding inhibitor) or to test or screen for protein folding inhibitor activity in a plurality of different test compounds, e.g., 10, 100, 1000, or more test compounds.

Preferably the determining involves comparing the ratio of folded protein to unfolded protein which occurs in the presence of the test compound to the ratio which occurs in the absence of the test compound.

The method can also involve a binding assay as an initial screen to identity test compounds which are more likely than others to be actual protein folding

inhibitors. This generally involves testing or screening a compound(s) or a library of molecules for binding to probe moieties, where the probe moieties may be or include peptides or proteins which correspond to sequence motifs present in a target protein. Compounds that bind to the probe moieties can be tested for their ability to inhibit the folding of biosynthetically-produced proteins, and thereby increase the amount or ratio of unfolded protein to folded protein. Typically folding inhibition prevents formation of active protein and/or enhances degradation and/or aggregation of the protein. Those skilled in the art are highly familiar with the constructioin and performance of the binding assays.

In preferred embodiments, the method involves contacting the protein with at least one chaperone protein. Such chaperone proteins bind to unfolded proteins or portions of proteins, and prevent folding or refolding. Thus, the chaperone protein(s) can be used, for example, to maintain a newly synthesized protein in an unfolded state, or to prevent a protein which has been unfolded from refolding. As a result, the chaperone proteins can be used prior to and/or during contact of a protein or peptide with a test compound. Thus, in some embodiments, the protein is contacted with the chaperone protein(s) prior to exposure of the protein to the presence of the test compound. Typically the test combination formed by the chaperone proteins and unfolded target protein is contacted with test compound, and the test combination is then subjected to conditions which tend to remove the chaperone protein from the unfolded protein. Target protein which has bound test compound will tend to remain at least partially unfolded. In addition, chaperones can be used in determining the ratio or amount of unfolded protein resulting from a test or control (for example, as described in U.S. Patent 5,679,582).

As recognized by those skilled in the art, different protein biosynthetic systems can be used, generally a system is selected which is appropriate for the target protein. Thus, in preferred embodiments, the protein biosynthetic system is an *in vitro* system, e.g., a eukaryotic protein biosynthetic system or a prokaryotic protein biosynthetic system. Particular systems are identified in the Description below and others are known to those skilled in the art and can also be utilized, and include the use of a biosynthetic system in which nascent polypeptides are held on the ribosomes for a time (e.g., as described in the Detailed Description).

Detection of unfolded protein, and thus determining whether a test

compound has increased the amount or the ratio of unfolded protein to folded protein, can utilize a variety of different techniques or approaches, including, for example, proteolysis with electrophoresis, aggregate sedimentation, binding to conformationally specific antibodies; binding to at least one chaperone protein, and determination of biological activity.

Many different types of test compounds can be utilized, whether in rational design, in random screening, or in a directed synthesis approach. Such test compounds can, for example, be multivalent binding compounds, complementary peptides, aptamers, and small molecules, e.g., small organic molecules, and members of a combinatorial library. The test compound can be or include a domain:domain interface sequence or a sub-domain:sub-domain interface sequence of the target protein.

Additional embodiments are described in the Detailed Description below.

In another aspect, the invention provides an assay method utilizing chaperone proteins to maintain a target protein in an unfolded (at least partially) state. The method involves binding an unfolded protein with at least one chaperone protein to form a test combination, contacting the test combination under non-denaturing conditions with a test compound, releasing the at least one chaperone protein, and determining whether the test compound increases the ratio of unfolded protein to folded protein, where an increase in that ratio indicates that the test compound is a protein folding inhibitor.

As indicated above, chaperone proteins can be used to prevent folding of a newly synthesized protein, or to prevent refolding of a previously synthesized protein which has been unfolded. Thus, in preferred embodiments, the unfolded protein is a previously synthesized protein subjected to unfolding conditions, or the unfolded protein is a newly synthesized protein, e.g., the at least one chaperone protein is present in a protein synthetic system during synthesis of the unfolded protein.

The stabilization of the binding of the chaperone proteins to unfolded protein can assist in the method, for example, by enabling the contact with test compound to be carried out at a later time. Thus, preferably the method also involves exposing the test combination to chaperone-binding stabilization conditions, e.g., chelating or otherwise removing contact with Mg ions.

In preferred embodiments, this method involves controls, ratio comparison, unfolded protein detection methods, test compound selection, and/or target protein or peptide selection as in the first aspect above.

In another aspect, the invention provides a method for identifying a protein folding inhibitor, where the method involves providing a peptide, where the peptide is a potential protein-stabilizing peptide and does not require unfolding, contacting the peptide with a test compound under non-denaturing conditions, and determining whether the test compound binds to the peptide. Binding of the test compound to the peptide indicates that the test compound is a protein folding inhibitor.

Preferably the peptide is selected to be one which is likely to be important in proper protein folding and/or one for which binding of another compound will disrupt the folding process and/or the final folded state. Thus, preferably the peptide can, for example, include (or be) a domain:domain or sub-domain:sub-domain interface sequence.

In preferred embodiments, the peptide is an isolated peptide or is in an intact protein or in a portion of a protein which includes at least two domains or sub-domains.

A target site can, in some cases, be accessible in a folded protein or polypeptide, for example, a site at an edge of a domain:domain or sub-domain:sub-domain interface. Thus, in preferred embodiments the peptide is in a folded polypeptide or protein.

The binding assay can be carried out in many different formats and arrangements, as is recognized by those skilled in the art. Such persons can readily select a variety of such test formats and detection techniques as appropriate for particular probe (target) peptides and/or test compounds and/or labels. In preferred embodiments, the test compound or a plurality of such test compounds are attached to a solid phase support. Alternatively, the peptide or a plurality of peptides is/are attached to a solid phase support.

As indicated, detection of binding can be carried out in a variety of ways. In preferred embodiments, the binding detection involves detection of a label attached to the test compound(s), detection of a label attached to a molecule containing the peptide, binding of an antibody (which may be labeled) to the peptide or protein, an electrophoretic mobility shift assay, or gel filtration.

In preferred embodiments, the method can involve the use of chaperones, selection of test compound, and determination of unfolding as in the first aspect above.

The method can also include the use of a direct protein folding inhibition assay, for example, as described in the above aspects. Such a direct protein folding inhibition assay can, for example, provide confirmatory indication in addition to the binding assay results.

In some cases, a protein folding inhibitor can bind to folded protein and disrupt that folding. Thus, in another aspect, the invention provides a method for identifying a protein folding inhibitor, where the method involves contacting a folded protein or polypeptide with a test compound under non-denaturing conditions, and determining whether the amount of unfolded protein or polypeptide is increased in the presence of the test compound.

In preferred embodiments, the test compound is selected to bind to a site that is buried in the folded protein structure. In such cases, a test compound which is a folding inhibitor can shift the natural equilibrium between folded and unfolded forms of the protein toward unfolding. Also in preferred embodiments, a test compound is selected to bind to a site on the surface of a folded protein. In such cases, a test compound which is a folding inhibitor can disrupt folding by altering the environment of the protein, for example, by distorting the protein folding structure and/or by altering the electrostatic or hydrophobicity characteristics of the local environment.

As particular sites in a protein are more likely than others to be accessible in a folded protein or polypeptide, in preferred embodiments, the test compound is selected to bind at a domain:domain or sub-domain:sub-domain interface.

As indicated above for detection of unfolded protein, such detection can involve a variety of different techniques or approaches. In preferred embodiments, the detection involves using proteolysis with electrophoresis, aggregate sedimentation, binding to conformationally specific antibodies; binding to at least one chaperone protein, or determination of biological activity.

In preferred embodiments, the method involves selection of test compound as described for the first aspect above.

In preferred embodiments, the method also involves a second protein folding

inhibition assay, for example as described in the first two aspects above, preferably an assay including a protein biosynthetic system.

It was further found that certain protein folding inhibitors could bind to probe peptides or proteins under denaturing conditions. Such an assay can be advantageous, for example, in allowing use of conditions in which the peptide or protein remains soluble, when it would not be sufficiently soluble under non-denaturing conditions. Thus, in another aspect the invention provides a method for identifying a protein folding inhibitor, where the method involves contacting a protein or polypeptide with a test compound under protein-denaturing conditions, where the protein or polypeptide includes a potential protein stabilizing peptide, and determining whether the test compound binds to the peptide. Binding of test compound to the peptide is indicative that the test compound is a protein folding inhibitor.

Preferred embodiments include the selections of peptide, test compound, methods of detecting binding, use of solid phase supports, and inclusion of additional confirmatory tests as described above. Thus, in particular embodiments, the peptide includes a domain:domain interface sequence; the peptide is in an intact protein or in a portion of a protein which includes at least two domains; the test compound or a plurality of test compounds are bound to a solid phase support; the peptide is attached to a solid phase support; binding is detected using a method selected from; detection of a label attached to the test compound, detection of a label attached to a molecule which includes the peptide, and binding of an antibody to the peptide; and the test compound is a multi-valent binding compound, a complementary peptide, a small molecule, or a compound including a domain:domain interface sequence.

The invention also involves the use of computers to identify potential protein folding inhibitors. Thus, in another aspect, a method for identifying a putative protein folding inhibitor is provided, where the method involves obtaining 3-D structural coordinates of a peptide or a plurality of peptides which form a structural domain or sub-domain of a protein (the coordinate description need not be limited to the domain or sub-domain), identifying a surface of the domain or sub-domain where the surface preferably forms an interface with one or more other structural domains or sub-domains, and docking a plurality of molecular structures (e.g.,

iteratively docking each member of a set of molecular structures) on the surface to determine the quality of fit. Identification of a molecular structure with a good quality of fit to the surface indicates that a compound with that molecular structure is a protein folding inhibitor. Generally the docking also involves describing an image or images or a set of negative or positive images or both of the surface, and docking the molecular structures on a surface image or images.

As inhibitors targeting N-terminal portions of a protein are preferred, in preferred embodiments the domain is an N-terminal domain of the protein.

The docking procedure can utilize a number of different algorithms. In preferred embodiments, the docking involves determining geometric constrains on fit of a potential binding compound and/or determining the putative molecular interactions of each molecular structure with the surface using computer calculation of the expected interaction free energy of the molecular structure with the surface. In preferred embodiments, the surface is identified using an implementation of a DOCK program or a modification or derivative of such a program.

The description of the molecular structure(s) can be performed with various programs available for such purposes. In preferred embodiments, the molecular structure or structures is described using an implementation of a CONCORD program or a modification or derivative of such a program.

The molecular structures can be structures of compounds which have been previously synthesized or can be virtual compounds. Thus, in preferred embodiments, the molecular structures are structures from a virtual compound library, a real compound library, or both.

The docking can be performed using selected sites on the surface or can be performed by walking structures over the surface. In preferred embodiments, the method involves selecting a site or sites on the surface for docking.

As the process can be performed for more than one surface in a protein, in preferred embodiments the method involves identifying a plurality of surfaces, which are preferably for a plurality of domains or subdomains of the protein.

As indicated in the Description below, domains and sub-domains in a protein can be identified in several different ways. In preferred embodiments the domain or sub-domain is identified using 3-D coordinates for the protein or a portion of the protein including the domain or sub-domain.

The computer identification of putative protein folding inhibitors can be further tested or confirmed using binding and/or protein folding assays, e.g., as described herein. In preferred embodiments, the method also includes testing the ability of a putative protein folding inhibitor to inhibit protein folding utilizing an assay which utilizes a protein biosynthetic system or a chaperone protein or both.

As with other methods of identifying protein folding inhibitors, molecular structures of a number of different types of test compounds can be used, for example, peptides, aptamers, and small molecules. The test compounds may be members of a combinatorial library.

In a related aspect the invention provides an additional method for identifying a putative protein binding compound. This method involves docking a plurality of molecular structures (e.g., iteratively docking each member of a set of molecular structures) from a virtual compound library on a surface or site of a protein to judge the quality of fit for each molecular structure, and choosing putative binding compounds by selecting molecular structures predicted to have good quality of fit. Preferably the virtual library is a virtual combinatorial library. The process can also involve describing an image(s) or set of negative or positive images or both of a protein surface, and using the image or images in the docking process.

Preferred embodiments include embodiments which involve: selecting a site or sites on the surface for the docking of a potential binding compound; where docking involves determining the putative molecular interaction of each of the molecular structure with the surface using computer calculation of the expected interaction free energy of the molecular structure with the surface; where docking involves the use of an implementation of a DOCK program or a modification or derivative of such a program; the molecular structure, or structures, is described using an implementation of a CONCORD program or a modification or derivative of such a program; identifying surfaces for a plurality of domains or sub-domains of the protein; the domain or sub-domain is identified using 3-D coordinates for the protein or a portion of the protein including the domain or sub-domain; providing at least one compound corresponding to a putative protein binding compound and testing the putative protein binding compound to determine whether that compound binds to the protein; providing at least one compound corresponding to a putative protein binding compound and testing that compound to determine whether the

compound alters a cellular property of the protein, e.g., a cellular property such as degradation rate, ligand binding, and biological activity; and determining the ability of at least one putative protein binding compound to inhibit protein folding utilizing an assay which utilizes a protein biosynthetic system or a chaperone protein or both.

Preferred embodiments include embodiments described for the preceding aspect.

The invention also provides methods for using protein folding inhibitors. In one aspect, the invention provides a method for inhibiting the cellular action of a protein, involving contacting the protein in a cell with a protein folding inhibitor active on the protein, preferably where the inhibitor specifically inhibits *de novo* folding.

In the context of this invention, the phrase "specifically inhibits *de novo* folding" distinguishes the inhibitors from compounds which inhibit folding of a previously synthesized protein following fast collapse of the protein from an unfolded state by changing conditions from unfolding conditions. Thus, such inhibitors target sites, for example, which are inaccessible following such fast collapse, or which are involved in irreversible folding.

Thus, in preferred embodiments, the inhibitor inhibits irreversible folding of the protein. Also in preferred embodiments, the inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, where the protein is unfolded from a folded state.

The term "cellular action", in reference to the function or functions of a target protein, means a biochemical activity or biochemical pathway in which the particular protein participates.

Various types of compounds can be used as inhibitors, including, for example, a binding peptide, a small molecule, a multi-valent binding compound, an aptamer, and an antibody, which may be a conformationally specific antibody. However, an antibody is less preferred in view of, for example, the difficulties of delivery and/or limitations on access to binding sites due to the large size of the molecule.

Preferably, the contacting is carried out *in vivo* in an organism, more preferably a mammal, and most preferably a human.

Also in preferred embodiments, the method is used in conjunction with

conditions which favor protein unfolding *in vivo*. Thus, for example, the contacting can be carried out in conjunction with heat shock treatment of an animal, preferably a mammal, most preferably a human.

The invention also provides a method for modulating a cellular process. The method involves contacting cells involved in or able to perform the process with a protein folding inhibitor active on a protein involved in that process. The inhibitor is one which specifically inhibits *de novo* folding.

The term "cellular process" refers to a biochemical, physical, and/or biological process which is performed in or by a cell or cells using cellular components, but distinguished from processes performed by a complex organism as a whole (e.g., walking; excretion of food wastes from a digestive tract). Preferably a cellular process is a multi-step and/or multi-reaction process.

The term "modulating" refers to changing the rate and/or extent of a process. Thus, in terms of a cellular process, the change can be an increase or a decrease in the rate and/or extent. The modulation can, for example, occur through reducing the amount of active protein produced, reducing the lifetime of active protein, and/or reducing the specific activity of a protein.

A "cellular property" of a protein means a physical and/or biochemical characteristic of the protein, or the behavior or response of the protein in a cellular process, e.g., degradation rate, half-life, ligand binding, biological activity, and receptor binding.

Preferred embodiments include those as described for the preceding aspect.

Also in preferred embodiments the modulating a cellular process involves or is enhancing the immunogenicity of a peptide or protein. As described herein, such enhancement can result from increased presentation of antigenic peptides due to increased degradation of unfolded proteins containing those peptides.

In a related aspect, the invention provides a method for modulating growth or proliferation of a cell. The method involves contacting the cell with a protein folding inhibitor active on a protein required for or regulatory of an essential cellular function. Preferably the inhibitor specifically inhibits *de novo* folding.

The terms "growth" and "proliferation" have their usual biological meanings in connection with cellular development. Thus, "growth" includes increase in cell size and/or numbers. "Proliferation" refers to the process and/or result of the

production of progeny cells from a parent or progenitor cell or cells, usually involving an increase in cell numbers of a particular type or types of progeny cells. The process may, in at least some stages involve changes in cell size and/or morphology and/or other cellular characteristics.

Preferred embodiments include embodiments as described for the preceding two aspect.

In another related aspect, the invention provides a method for treating a disease or condition in a patient by administering a therapeutically effective amount of a protein folding inhibitor to the patient. The inhibitor preferably specifically inhibits *de novo* folding of a protein involved in the disease or condition.

A "therapeutically effective amount" is an amount sufficient to reduce at least one symptom of a disease or condition, which can include reducing the severity of the disease or condition, cure the disease or condition, reduce a deleterious effect of the disease or condition, or produce other clinically relevant effect.

A protein is "involved in the disease or condition" if the protein is important in creation, maintenance, or development of the disease or condition. Thus, for example, the protein may be an essential protein of an infecting organism, a protein important in the pathogenesis of an infecting organism, a protein of a host important for the establishment, development, or course of an infection, and a host protein which directly or indirectly contributes to the initiation, development, course, or effects of a disease or condition (for example, a protein produced in excess amount can cause a condition which can be alleviated by reducing the amount of active protein or the activity associated with the protein, such as with a folding inhibitor, a competitive inhibitor, an antagonist, or an expression inhibitor). Those skilled in the art are familiar with the characterization of proteins involved in diseases and conditions and can readily recognize such proteins.

Preferred embodiments include embodiments as described for methods of using protein folding inhibitors in preceding aspects.

The invention also concerns pharmaceutical compositions which include a protein folding inhibitor. Preferably the inhibitor specifically inhibits *de novo* folding. The composition may also include a pharmaceutically acceptable carrier or excipient. Generally the composition will be or will be made to be sterile, meaning that the composition is sufficiently free of viable organisms which are not intended

to be part of the composition to satisfy United States regulatory requirements for administration of a composition to a mammal, preferably to a human. Preferably the inhibitor is prepared in a manner which will satisfy United States regulatory requirements for drugs to be administered to humans.

In preferred embodiments, the inhibitor is as described for preceding aspects (including types of compounds and/or target sites or peptides). Thus, in preferred embodiments the inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, wherein said protein is unfolded from a folded state; the inhibitor inhibits irreversible folding of the protein; the inhibitor is a multi-valent binding compound, a binding peptide, an aptamer, or a small molecule.

In the methods of using a protein folding inhibitory, the effects may be regulated or titrated. For example, the effects can be regulated by regulating the amount of inhibitor utilized and/or the inhibitor may be selected to have a desired level of activity. In such cases, the inhibitor may inhibit folding of only a portion of the target proteins and/or reduce the activity to some degree which may be less than complete inhibition.

In the methods of using protein folding inhibitors and the pharmaceutical compositions described herein, preferably the inhibitor is of a type as described herein, or which would be identified and/or produced by the identification and/or production methods described herein.

Likewise, the invention provides a method for making a pharmaceutical composition. The method involves screening to identify a protein folding inhibitor, where the screening involves the use of a protein biosynthetic system assay; and synthesizing the compound in an amount sufficient to provide a therapeutic response when administered to an individual suffering from a disease or condition involving the target of the inhibitor. Preferably the inhibitor is one which inhibits *de novo* folding of the protein.

Those skilled in the art are familiar with methods of synthesis appropriate for different types of compounds, including biosynthesis, total chemical synthesis, and chemical modification of existing compounds. Thus, appropriate synthesis and/or purification methods can be readily selected and/or designed following identification of a particular inhibitor compound.

The screening can involve any of the identification methods described herein.

As indicated above, in aspects of the present invention involving identification of protein folding inhibitors, a library of molecules can be used. Generally the molecules in the library will be screened for binding to probe moieties that may be peptides or proteins which correspond to sequence motifs present in a target protein. Compounds that bind to the probe moieties are preferably tested for their ability to inhibit the folding of biosynthetically produced proteins and thereby prevent formation of active proteins.

In preferred embodiments, the molecules are in libraries or collections of such molecules made by conventional chemical synthetic methods. In such an embodiment, the compounds can be synthesized in a combinatorial chemical library, preferably the chemicals are small molecules, more preferably small organic molecules, which may be members of a combinatorial library. In some embodiments, the molecules are combinatorial libraries of peptides synthesized on polymer beads or the like. In some embodiments, the molecules are libraries of complementary peptides. In some embodiment, protein folding inhibitors are identified by computer-aided computational methods that scan existing, or theoretical, chemical structures and compare their shapes to the known or computed surfaces of target proteins. In some embodiments of the present invention, folding inhibitors are selected from or identified in a library of aptamers. "Aptamers" are RNA molecules, which may contain modified nucleic acids, which bind specifically to particular target peptides.

Folding inhibitors identified by the method of the present invention are, commonly but not necessarily, conformationally rigid molecules or molecules that contain a conformationally rigid scaffold, which contain sufficient hydrophobic chemical groups so as to facilitate penetration of cell membranes. Conformationally rigid molecules have an increased affinity for a target peptide because their binding to the generally flexible unfolded protein is entropically favorable. Examples of inhibitor types which are not conformationally rigid include with the exception of linear or branched peptides.

For initial screening, advantageously probe moieties used to test folding inhibitors of the present invention are preferably peptide sequences from

tetrapeptides to hexadecapeptides (i.e., 4-16 amino acid residues). Preferably such probe peptides correspond to loop regions contained within the hydrophobic core of a native protein. Probe peptides may also correspond to a sequence associated with biological activity. In other embodiments of the invention, sequences having a distinguishing mutation will be probe sequences. Yet other probe peptides correspond to sequences which, in the native protein, form beta strand or sheet motifs. In general, probes will contain sequences found in the holo-protein, whose binding results in inhibition of folding resulting from steric hindrance, inhibition of protein strand packing, disruption of electric charge and/or cross-linking of polypeptide strands. In still further embodiments of the invention, a reporter, for example a dye molecule, is added to the probe. Probe moieties may consist of a target sequence or a slightly larger peptide containing the target sequence as a probe. In further embodiments of the invention, polar or charged chemical groups or a soluble second polypeptide may, in some instances, be added to the probe moiety to improve its water solubility where this is beneficial to perform binding assays in aqueous solution where such assays are desirable.

Compounds identified as exhibiting binding to a probe moiety are preferably tested for specificity and affinity of binding to the target peptide. In certain embodiments, such testing involves on-bead binding assays by two probes, each containing a different dye moiety. In some embodiments, such testing involves binding of a probe to each of several different peptides. In some embodiments, such testing involves binding to a random combinatorial library of peptides.

In preferred embodiments of the invention, initial low-affinity leads, identified by application of the method of the present invention, will guide repeated cycles of synthesis, e.g., combinatorial synthesis, and screening, leading to higher affinity binding compounds. Additionally, dimerization, e.g. chemical linking of two or more identified binding-compounds with a chemical linker, will yield higher affinity binding compounds. In such chemically-linked dimers, the monomers may be identical to one another or may be different.

For example, in some embodiments, a dimer may consist of identical peptide binding monomers whose target is repeated in the protein structure, or the dimer may consist of two different monomers, each binding to a unique sequence within the target protein. In certain embodiments of the present invention, the two target

sequences within the target protein may be in different domains in multidomain proteins that fold post-translationally, e.g., in bacterial multidomain proteins. In yet other embodiments of the invention, for eukaryotic proteins, the target sites will be in different sequences in the same protein domain.

In some embodiments, sites may be selected within different domains in a eukaryotic protein, or may include a site located in the sequence between two domains, where the protein is a mutant resulting in the fusion of at least two proteins or domains. Such fusions can, for example, occur in certain oncogenic proteins.

Such dimers are preferably tested for specificity and affinity of binding to the folded and unfolded form of the target protein. Such testing is preferably done in cell-free protein-translation systems, where binding of the lead compound is assayed for its ability to inhibit folding of the target protein in the context of translation.

In certain embodiments of the invention, such a folding inhibition assay is performed by limited proteolysis of a translation product. The limited-proteolysis assay indicates folding inhibition because unfolded or incompletely folded proteins are more susceptible to proteolytic digestion.

In other embodiments of the invention a folding inhibition assay involves detection of protein aggregation. The protein-aggregation assay indicates folding inhibition because incompletely folded proteins are more susceptible to aggregation.

In yet other embodiments of the invention, binding of a conformationally specific antibody is assessed to determine the proportion of protein that has folded into the native conformation. In still other embodiments of the invention, the biological activity of the protein is determined to assess the proportion of the protein that has activity.

In further embodiments of the invention, testing is done *in vitro,* where the lead compound is assayed for its ability to inhibit refolding of the holo-protein. In further embodiments of the invention, compounds are tested in biological assays on cultured cells or *in vivo*, and the lead compound is assayed for its ability to inhibit folding in the context of biosynthesis or refolding in the context of stress-induced unfolding or in another relevant biological assay.

Aspects utilizing protein folding inhibitors make use of active compounds of the types indicated above, or compounds resulting from such identification methods as described above.

By the term "protein biosynthetic system" is meant a system including but not limited to ribosomes (either eukaryotic or prokaryotic) that is capable of synthesizing a protein utilizing either RNA or DNA or both as a template. Thus, where an appropriate template is provided under protein synthesis conditions, the system will synthesize the encoded protein or polypeptide. Typically the system is an *in vitro*, cell-free system, but in some cases may utilize intact cells, either in a cell culture or *in vivo* in a complex organism.

What is meant by "protein synthesis conditions" is conditions of a biosynthetic system that are conducive to or stimulate protein synthesis.

The term "unfolded protein" refers to a polypeptide which does not possess stable or native tertiary structure, generally referring to a polypeptide that under certain conditions possesses stable, native tertiary structure, and/or a polypeptide which does not possess secondary structure that under certain conditions possesses native secondary structure, and/or a polypeptide or protein which does not possess native structures such as secondary, tertiary or quaternary structure where a native structure for the polypeptide or protein does possess one or more such structures and/or a protein or polypeptide that does not possess a stable three dimensional structure formed by the polypeptide chain, under given conditions but which possesses such a structure under other conditions.

The term "folded protein" refers to a polypeptide or protein that possesses stable, native tertiary structure or other structural determinants or three dimensional structure characteristic of its native form.

A "chaperone protein" is a protein that functions to prevent or lessen the tendency of other proteins to form misfolded aggregates upon folding and/or which facilitate folding of said proteins. Chaperones include, but are not limited to, members of the Hsp70, Hsp60, Hsp90, and Hsp40 classes. Thus, the term "at least one chaperone protein" refers to one or a plurality of different chaperone proteins (e.g., dnaK, dnaJ). Generally such chaperones will bind to an unfolded peptide or polypeptide by virtue of the peptide's unfoldedness or exposure of hydrophobic amino acids to solvent. The binding of peptide may be stabilized by chelation of magnesium in the solvent medium or by the absence of sufficient concentrations of adenosine triphosphate (ATP).

In the context of protein biosynthetic systems, the term "*in vitro* system" refers to a protein biosynthetic system lacking intact cells; in the context of a cellular biosynthetic system, the term indicates that the system includes living cells that are maintained outside of the body of a complex organism; in the context of an *in vitro* folding assay, the term indicates that the system includes protein molecules, preferably purified protein molecules, without a protein biosynthetic system and without living cells.

A "eukaryotic protein biosynthetic system" is a protein biosynthetic system in which functional ribosomes have been derived from a eukaryote.

Similarly, a "prokaryotic protein biosynthetic system" is a protein biosynthetic system in which functional ribosomes have been derived from a prokaryote.

The term "proteolysis" refers to digestion or degradation of a protein, polypeptide, or peptide, e.g., complete digestion of a protein to amino acids and/or short peptides (proteolysis) or limited proteolysis which is the digestion of a folded or unfolded protein to characteristic fragments or in some cases to amino acids and short peptides. Generally such digestion is performed by one or more proteases.

In connection with polypeptides, the term "aggregate" refers to a non-specific association of one or more types of polypeptides to form, for example, clumps or particles. Thus, the term "aggregate sedimentation" refers to a method of concentrating protein aggregates by a sedimentation method, e.g., by centrifugation or gravitational sedimentation.

The term "antibody" includes the usual biological meaning in referring to proteins as produced by a vertebrate immune system, but also including portions of such molecules which retain antigen-binding activity, and can also include chemically modified derivatives which retain antigen-binding activity.

The term "conformationally specific antibodies" refers to antibodies whose binding to a protein antigen differs based on the conformation of the protein antigen. Thus, for example, a conformationally specific antibody binds differentially to corresponding protein antigens in folded, partially folded and unfolded states. The term can include antibodies that bind to protein epitopes constructed of protein tertiary or secondary structure. Such antibodies can, for example be used in assays

in the present invention where specific binding or lack of binding are representative of the informative result of an assay.

What is meant by "determination of biological activity" is an assay or method which identifies the quality, and usually but not always quantity of level of: the biochemical activity of a protein, a biological process, a cellular effect, or effect of a living organism or part of the organism. Typically, the activity is a normal cellular activity or an activity associated with a cancer, e.g., tumor growth.

The term "bi-valent binding compound" refers to a ligand having two binding valencies, each of which binds to a peptide (which may be the same or different) within a single protein or polypeptide, where those peptides are separated from one another, typically separated in the primary sequence of a protein. Preferably the respective bindings can occur contemporaneously. Similarly, the term "multi-valent binding compounds" refers to ligands which have a plurality of such binding valencies, e.g., 2,3,4, or even more.

The terms "complementary peptide" and "aptamer" are described in the Detailed Description below. In addition, "complementary peptide" refers to peptide containing a sequence of amino acids which, in general, possess opposite hydropathies to the amino acids in the sequence of another peptide. "Aptamer" refers to an RNA-containing (and/or modified nucleotide-containing) molecule, where the molecule or at least the nucleotide portion binds to a particular target peptide.

The term "small molecule" refers to a chemical compound, preferably an organic chemical compound comparable in molecular mass to common drug molecules; and thus has a molecular mass of less than 3000 Daltons, preferably less than 2000 Daltons, still more preferably less than 1500, 1200, or 1000 Daltons, and most preferably less than 800, 600, or 400 Daltons. Usually, but not necessarily a small molecule of the present invention will be in the range of 300Da to 1200Da. Furthermore, such small molecules will often be elements of a combinatorial chemical library or other chemical library or archive.

A "protein stabilizing peptide" is a peptide located within a larger polypeptide or protein that contributes stability to the folded structure of said polypeptide or protein. Preferably the peptide is one which preferably assumes a conformation which contributes to proper folding or maintenance of the folded form,

or which preferentially interacts with another peptide or peptides in the polypeptide or protein in a manner which contributes to formation or maintenance of the folded form.

The phrase, "does not require unfolding", indicates that a polypeptide or peptide probe (which may correspond to a protein stabilizing peptide) remains soluble and potentially active in a binding assay or other relevant assay under non-denaturing conditions. Generally this indicates that a target site or sequence is accessible to test compounds without denaturation. The polypeptide or peptide, if requiring unfolding, would require exposure to sodium dodecyl sulfate or other denaturant as a means of maintaining solubility and binding activity in such an assay.

The term "subdomain:subdomain" interface refers to a peptide sequence or sequences of a protein, where the peptide is located at an interface adjoining at least two structural subdomains, each consisting of protein secondary structural motifs.

"Protein denaturing conditions" refers to solvent, temperature and/or pressure conditions which favor protein unfolding.

What is meant by "non-denaturing conditions" is solvent conditions which favor protein folding or maintenance of the folded state of a protein.

In the context of this invention, the term "bind" refers to the formation of non-covalent chemical bonds and/or other energetically favorable interactions (e.g., hydrogen bonds, van der Waals attraction, electrostatic attraction) between a test compound and a peptide contained in or derived from a protein of interest (or to an unfolded form of said protein), where the bonding is of a specific nature and is energetically favorable. Generally the bonding is specific to a particular sequence or binding site geometry. In certain cases, the binding may also include formation of chemical bonds, i.e., covalent bonds.

A "domain:domain interface sequence" means a peptide sequence of a protein, where the peptide is located at an interface adjoining at least two structural domains.

What is meant by "intact protein" is a holoprotein.

What is meant by "solid phase support" is a solid material such as a microbead composed of polystyrene, polyethylene glycol-grafted polystyrene or some other material or to a plastic or other solid surface.

A "solid phase support" indicates a solid material, for example, a microbead composed of polystyrene, polyethylene glycol-grafted polystyrene or some other material, or a plastic or other solid surface, to which second material is or may be attached, directly or indirectly. Thus, for example, a solid support may have attached to it a test compound or compounds, other ligands, or peptide or protein probes.

What is meant by "detection of a label" is the identification of the presence of a label moiety. The detection can, for example, involve spectroscopic, spectrophotometric, fluorometric, colorimetric, or other methods or labels as understood by those skilled in this art. Preferably the detection involves visual detection of a dye (label), wherein the dye has colored a solid support by virtue of the dye being bound covalently or non-covalently to a molecule that has bound specifically to the support, or where the molecule has bound to a second molecule already bound to the support. The dye may be of a fluorescent nature. In other embodiments, a label is not be a dye, but rather may be a radionuclide, where detection may, for example, involve exposure of a photosensitive surface, or other method of detecting the emission of particles, e.g., nuclear particles or other radiation.

In a multi-domain protein, e.g., where the sequence of amino acids of the whole protein or polypeptide can be ordered by consecutive integers ranging from 1 to n (where n>1) or from k to n (where n>k) beginning at the N-terminal end of the polypeptide, a protein domain is said to be an "N-terminal domain" relative to a second domain or a plurality of domains if the N-terminal domain is composed of a polypeptide chain or chains such that some or all of the amino acids making up the domain are ordered by integers that are smaller than any of the integers that order the amino acids in the second domain or other of the plurality of domains. For the methods of this invention, preferably, but not necessarily, an N-terminal domain is located fully or primarily in the ½ of the protein or polypeptide nearest the N-terminal end of the mature protein or polypeptide.

A "putative molecular interaction" refers to a prediction of binding between a protein (or protein domain or receptor) and a second molecule, e.g., a small molecule, where the prediction of binding is based on a computer algorithm that indicates molecular complementarity between the protein (or receptor) and the

second molecule, and/or a prediction based on an indication of a good quality of fit derived from other parameters such, but not limited to minimization of free energy. Thus, a "putative protein folding inhibitor" is a compound or molecular structure predicted by computer-based or other representational methods to bind to a target protein, generally at a location which is expected to result in folding inhibition.

A "virtual compound library" refers to an electronic library of molecular structures, e.g., where the molecular structures represent those structures deduced for the products of a combinatorial chemical synthesis protocol. Thus, an electronic library which includes the molecular structures of all or some of the compounds that would be predicted to be produced in a combinatorial chemical library or set of such libraries where the members are predicted for the synthesis protocol. Such compounds do not have to be produced in physical form to construct the virtual library. Such libraries may, for example, contain 10^2, 10^4, 10^6, 10^8, 10^9, or more such compounds.

The phrase "negative or positive images" refers to contour maps (an image surface or volume) corresponding to the surface of a protein, protein domain, or domain:domain interface (or subdomain:dubdomain interface), where the negative image is produced by a computer algorithm which fills pockets, clefts, and other invaginations of the surface, e.g., with sets of spheres, to define a contour map. The centers of the spheres form the negative surface. The positive image is produced by a computer algorithm creates a contour map of the surface, e.g., by placing spheres inside the molecular surface of the protein or domain or inside a potential ligand. In such cases, a positive image is formed by the centers of the spheres which form bumps or other raised structures on the surfaces. Other computational methods instead of a sphere filling approach may also be used to define the contour map or image surface.

The term "docking" refers to a computer-based process which determines that a potential ligand, placed in an electronic representation within a site made up of part of a protein, or domain, or subdomain surface in a variety of orientations, does or does not, in at least one orientation, enter the excluded volume of the protein. Additionally, the docking determines the contacts between the ligand and receptor. Docking may also include calculation of factors such as electrostatics, molecular

mechanics, buried surface area, packing, and free energy of solvation to determine goodness of fit.

In the context of computer-based methods of this invention, the term "molecular structure" refers to the 3-dimensional atomic coordinates of a compound, e.g., a test compound or ligand. The coordinates can be used, for example, in producing a positive and/or negative image of the compound, and further in a docking process.

The term "quality of fit" refers to a measure of complementarity of shape between two compounds, e.g., between a test compound or ligand and a receptor or similar site on the surface of a protein, protein domain, or subdomain. A "good quality of fit" is one which indicates sufficient complementarity between the compounds so that a test compound is judged a putative binding compound or ligand based on the likelihood of binding. The threshold criteria for such identification can, for example, be set based on comparison of quality of fit with known binding pairs and compounds known to bind poorly or not at all to particular sites.

In the context of computer software used or useful for this invention, the term "modification" refers to an advanced, improved, or otherwise changed version of the software, e.g., of the program DOCK or CONCORD. Thus, the modification will include the analysis of the prior version, but can include extensions or refinements of that analysis. Usually, a modification will be a later version which extends the analysis options, improves display capabilities, corrects defects or bugs, or improves or adds to the parameters utilized in the analysis.

Also in the context of computer software used or useful in this invention, the term "derivative" refers to computer software which incorporates important elements of a previous program. In particular a "derivative" of a program, e.g., DOCK or CONCORD, which can predict the likelihood that a molecular structure will bind a site or have a good quality of fit on a protein, domain, or sub-domain surface and/or provide a computer representation of a molecular structure will utilize important elements of a prior program which can also provide such prediction and/or provide such structures.

By "comprising" is meant including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or

may not be present. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

Other features and embodiments will be apparent from the following description of preferred embodiments and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows space fill and backbone diagrams for PGHS-1 showing the locations of a peptide distinguishing PGHS-1 and PGHS-2 and a peptide common to both proteins. The two structures at the top are spacefill diagrams (different views) showing amino acids 256-262 (peptide distinguishes PGHS-1 and PGHS-2). The structure at the bottom right is a backbone structure. Amino acids 300-306 are visible in the backbone diagram but not in the spacefill diagrams because they are buried in the folded protein. The two sets of peptides seen in the backbone diagram belong to the A and B chains of this homodimer.

Figure 2 shows a generalized structure of a peptidosteroidal bivalent protein folding inhibitor.

Figure 3 shows a generalized structure of a bivalent branched peptide protein folding inhibitor.

Figure 4 shows a generalized structure of a bivalent fusion peptide protein folding inhibitor.

Figure 5 shows a generalized structure of a cyclic fusion peptide protein folding inhibitor.

Figure 6 is a schematic structure of a cyclic peptide complementary to an H-ras loop peptide.

Figure 7 is a flow chart of an exemplary computer-based method for identifying protein folding inhibitors.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As was described in the Summary above, the present invention involves the identification and use of protein folding inhibitors, and presents a number of different methods for such identification and use. Additional features, embodiments, and examples are described herein.

For example, in addition to inhibiting folding of a protein in the context of its biosynthesis, folding inhibitors can also inhibit the refolding of a stress-unfolded protein by binding to the unfolded protein, and especially to polypeptide sequences that become exposed in proteins as a result of unfolding and when such proteins are bound to chaperones. Hence, folding inhibitors or the clinical and/or pharmacological inhibition of folding can also be viewed and used as an adjunct to heat shock (or other stress) therapy.

Folding inhibition can also be used against specific proteins to elicit and/or potentiate an immune response against the protein target. This occurs because the protein that is inhibited from folding is degraded and presented (in vertebrates) by class 1 MHC molecules to T cells. Rapid or enhanced degradation results in quantitatively greater antigen presentation, potentiating a CD8 T cell response where the presented peptide is antigenic.

The following description concerns examples of methods and strategies pertaining to: 1) Choice of target; 2) Synthesis of probes (where applicable); 3) Synthesis of combinatorial library (where applicable); 4) Screening of library for lead compounds (where applicable); 5) Design of lead compounds (where applicable); and 6) Folding assays. The description herein provides further embodiments, e.g., methods for identifying, testing, optimizing, and using protein folding inhibitors, as well as types of inhibition compounds and types and examples of target protein and peptides.

Folding Inhibitors that Bind to Unfolded Polypeptide Chains (Class 1).

Targets

A protein of medical and/or research interest is chosen. As examples, three candidate proteins are discussed: human H-ras (oncogenic mutation valine 12), HIV-1 reverse transcriptase, and human prostaglandin H_2 synthase (PGHS-2). Based on a knowledge of the amino acids sequence of the target, or parent protein (see National Center for Biotechnology Information Genbank), and in various instances on the three dimensional folded structure of the target protein, formulations termed "probes" are constructed. One class of probes consists of a formulation containing a peptide of between 4 and 16 amino acids in length which has been chosen from the known amino acid sequence of the target protein. A second class of probes contain either the entire amino acid sequence of the protein or a peptide larger than 16 amino acids. Such probes are termed "shotgun" below. Probe peptides can be chosen from the very large number of short peptides forming any polypeptide chain. However, the following constraints serve to optimize lead compound discovery.

Regarding peptides between 4 to 16 amino acids, in certain embodiments two non-overlapping sequences (with respect to their positions in the target protein) are chosen so that each contains hydrophobic amino acids and one or more charged or polar amino acid(s). Additionally, at least one of these sequences should distinguish the target protein from related proteins (if such distinction is desired at the pharmacological level), and sequences that are common to many diverse proteins should be avoided (unless a non-selective lead compound is desired).

Part of the subject of this invention concerns how to greatly enhance the affinities of peptide binding molecules for unfolded proteins. This can involve the synthesis of bi- or multivalent binding compounds (or receptors), and, as described herein, the novel biological targets proposed in this disclosure are able to take unique advantage of this multivalent targeting strategy. As a comparison, the antibody molecule represents a natural solution to the problem of increasing binding affinity through mutivalency. In situations where an antigen contains multiple occurrences of an epitope, such that each of the antibody's antigen binding sites may bind an epitope at the same time, the increase in binding energy of the whole antibody (over that of the individual binding sites) approaches the product of the binding energies of each individual antigen binding site.

This principle has been used to increase both the binding energy and selectivity of opioid receptor ligands and ligands of the acetylcholine receptor. In these cases, the ligands are bivalent; one valency binds at one site on the receptor and the other binds at a pharmacologically distinct site. However, the length and mobility of the chemical linker joining the two pharmacophores is critical because the binding sites are essentially fixed in space relative to one another. As will become clear in a later section this linker constraint occurs because the above targets are folded proteins. Some of the folding inhibitors that are the goals of the methods of this disclosure will also constitute bi- or multivalent entities, however, linker length and flexibility requirements will not be stringent (as above) because multiple target sites within an unfolded protein are not as spatially constrained relative to one another as are sites on surfaces of folded proteins; additionally there are more potential binding sites. As a result, bivalent or multivalent lead compounds that are addressed by the present invention are far more likely to bind the unfolded protein target than its folded counterpart.

Domain Structure

Where the domain structure (as defined by Richardson, J.S. Methods in Enzymology 115, 341-380 (1985)) of the target protein is known, both peptide probes should preferably be located in the same domain (the "domain rule"); this will maximize the affinity of bi-valent lead compounds (to be explained). Here binding of the bi-valent (or multivalent) lead compound to the protein occurs more readily (and with higher energy) when both binding sites are concomitantly accessible. This is guided by the fact that during biosynthesis a domain remains unfolded (i.e., its various core peptides concomitantly accessible) until synthesis is completed and the chain is sufficiently extruded from the ribosome. In contrast peptides located in different domains will not always be displayed concomitantly (e.g., if one is located in a domain core), and therefore will not always be accessible for binding at the same time. This occurs in eukaryotes because protein domains fold co-translationally (while ribosome-bound) and sequentially. Thus, an already completed domain may have folded, and rendered a core site inaccessible, prior to the availability of a second site in an adjacent domain.).

At least two exceptions to the "domain rule" occur. Probes can be chosen from different domains when: 1) one probe constitutes a solvent accessible (surface exposed) peptide in an N-terminal domain, and; 2) when multiple domains or parts of different domains remain unfolded (concomitantly) while ribosome-bound. In these cases probes may be taken from different domains because the two sequence are concomitantly available for binding. Choosing target peptides without considering domain structure can, nevertheless, result in the discovery of folding inhibitors, but this is a less preferred formulation of the present invention.

There are several methods by which the amino acid sequence of a protein domain can be derived or specified. For example, the boundaries of structural domains are defined in terms of their amino acid sequences in many publications of crystal and solution structures and can also be inferred based on sequence homology (where the term "homology" is used to refer to sequence "similarity") analysis as described below (i.e., a domain in a protein of unknown structure is defined as a domain by showing homology of its amino acid sequence to that of a known domain from a structurally defined protein) and/or from domain databases such as Prodom (http://protein.toulouse.inra.fr/prodom.html) which defines specific domain and subdomain regions for a very large set of proteins, based on sequence homologies. Additionally, the CATH database: (http://www.biochem.ucl.ac.uk/bsm/cath) contains all of the proteins of the Brookhaven Protein Database, listing most domains and their boundaries. The Three-Dee Database of Domain Definitions contains similar information. Any other similar databases can also be utilized.

Also, domain structure can be predicted for proteins whose structures have been solved but for whom clear domain definitions are lacking. This is accomplished, for example, by utilizing any of several computer programs; e.g., the program DOMAK (see, e.g., Siddiqui, A.S., Barton, G.J. Protein Science (1995), 4:872-884). Additionally, for those of skill in the art, the designation of a structural domain can generally be accomplished by visual inspection of a protein structure.

Alternatively, where the domain structure is unknown, the probe peptides should preferably be chosen so that they lie within 50 amino acids of one another in the protein sequence (to maximize the possibility that the target sites will lie within the same protein domain).

A Further Optimization in Target Sequence Choice

In a general form, the invention allows the target sequences to be chosen from any sites within the polypeptide sequences of a domain. However, a level of optimization is produced if at least one amino acid sequence (probe) is that of a sequence located wholly or partly within the hydrophobic core of the folded protein (i.e., a solvent inaccessible peptide). This can be judged, for example, by referring to the publication of the crystal or solution structure which designates solvent accessible and inaccessible regions of the polypeptide.

Alternatively the coordinates of the structure can be obtained, e.g., downloaded from the Brookhaven Protein Database (http://www.pdb.bnl.gov/browse_it.html), and opened in a molecular graphics program; e.g., Rasmol 2.5. (RasMol 2.5 Molecular Graphics Visualization tool. Roger Sayle, BioMolecular Structures Group, Glaxo Research & Development, Greenford, Middlesex, UK. June 1994). The (Rasmol) command "select buried" (or other similar command from a different program) will allow designation of all amino acids in the protein that have a tendency to be buried (not solvent exposed). Although some of these may actually be exposed to solvent, many will be buried and thus the command provides a strong indication of buried amino acids. A particular peptide is then selected (e.g., "select 8 -16") and colored (e.g., "color red") while the protein is "displayed" as "backbone." Then "all" is selected and displayed as "spacefill." A "buried peptide" will disappear from all views (perspectives) of the resulting structure (perspectives are manipulated by the scroll bars). See drawing 1. As indicated, buried amino acids can be identified and appropriate peptides can be selected using the appropriate commands in other molecular graphics programs.

For those proteins whose structures have not been directly solved, but which are homologs of proteins of known structure (e.g., PGHS-2 structure not published at time of this writing; but PGHS-1 structure was published); then the locations of peptides (and their relationship to structural domains) within the protein homolog lacking a published structure correspond to those of peptides in homologous regions of the protein homolog of known structure (e.g., the PGHS-1 peptide aa 33-43 and PGHS-2 peptide aa 19-29 are homologous regions as is readily ascertained by aligning the two proteins by amino acid sequence. (Similar relationships hold for paralogous proteins, where homologies occur for given domains only and not for the whole sequence.) This was done here by performing NCBI BLASTP (Blast 2) using

one sequence as the subject and the other as the query. (The BLASTP program allows comparison of homologous stretches of sequence between the two proteins by assigning one sequence as the "query" and the other as the "subject.") For instructions see National Center for Biotechnology Information website: http://www.ncbi.nlm.nih.gov. Other sequence alignment tools are readily available and can likewise be utilized.

For proteins of unknown structure that are not related to any proteins of known structure, one probe preferably contains at least 75% hydrophobic amino acids. A hydrophobicity plot of the amino acid sequence will define hydrophobic, and potentially buried, regions of the protein. Otherwise visual inspection of the amino acid sequence and designation of "polar," "charged," and "hydrophobic" to each amino acid is sufficient. Those skilled in the art are familiar with the hydrophobicity of various amino acids and with hydrophobicity plots.

EXAMPLES ILLUSTRATING CHOICE OF TARGET SEQUENCES:

H-ras

Sequence **1** consists of amino acids 5 to 16 and sequence **2** consists of amino acids 73 to 86: (The protein contains a single folded domain.) (Sequence and numbering in accordance with Barbacid, M. (1987) Ann. Rev. Biochem. 56:779-827; Brookhaven PDB site: http://www.pdb.bnl.gov/pdb-bin/send-pdb?id=521p)

1 KLVVVGA*V*GVGK (contains distinguishing mutation shown in italic)

2 RTGEGFLCVAINN

shotgun amino acid sequence 1 to 189

HIV-1 reverse transcriptase

This protein consists of an A and B chain which are identical over their N-terminal sequences. The target sequences chosen below are found in both chains and occur in domain 2, based on CATH classification. Sequence 1 consists of amino acids 95 to 101 and sequence **2** consists of amino acids 205 to 214: (For sequence and numbering see Ren, J. (1995) Nat. Struct. Biol. 2(4):293-302; Brookhaven PDB site: http://www.pdb.bnl.gov/pdb-bin/send-pdb?id=1rt1)

1 PHPAGLK

2 LRQHLLRWGL

shotgun amino acids 1 to 250

Neither peptide is buried in the folded protein based on visual inspection of an X-ray crystal structure model (Due to the architecture of folded HIV-1 reverse transcriptase, most peptides are not buried.). The peptides were chosen to increase the likelihood, however, that binding of a bivalent agent (described later) will be more likely to occur when the protein is in an unfolded state. This results from the peptides' relative positions (opposite sides) in the folded protein, which make concomitant binding to both positions in the folded protein less likely. A bivalent agent with affinities for two peptides (within a target protein) can easily bind those peptides when the protein target is unfolded and the polypeptide chain flexible;. However, in a folded target protein, binding of such a bivalent agent is constrained because of the length and conformation of the linker joining its two valencies and because the target peptides are in fixed positions or may not be surface exposed. This is beneficial because binding to the folded protein could (through competition) reduce the concentration of agent available for binding to unfolded protein. In most cases, however, even random choice of peptide targets within a domain will result in the tendency of a bivalent ligand to avoid the folded protein (stochastic). (Also, note that binding to a folded protein is not a predictor of biochemical or biological response, whereas binding to an unfolded protein is indicative of inhibition.)

Prostaglandin H_2 synthase (PGHS-2)

To derive target sequences for PGHS-2 (for sequence see Jones et al. (1993) J. Biol. Chem. 268(12):904), targets were chosen initially for PGHS-1 (see Figure 1) based on it having an available structure (see Dewitt & Smith (1988) Proc. Nat. Acad. Sci. USA 85:1412-1416; Brookhaven PDB site: http://www.pdb.bnl.gov/pdb-bin/send-pdb?id=1pth). Although this is not obligatory, it represents an optimization by identifying one sequence as "buried," based on the known structure of a protein homolog. One of the PGHS-1 target sequences, 256 to 262, was chosen because this peptide distinguishes PGHS-1 from PGHS-2. The second PGHS-1 sequence consists of amino acids 300 to 306 which is a buried peptide, common to both proteins. The strategy is to align the two proteins (based on sequence homology; see diagram 1 which does this by employing BLASTP) and pick two corresponding

peptides in PGHS-2. These peptides are predicted to occupy similar positions in each protein based on the sequence alignment. However, because of the uniqueness of peptide 256-262 (numbers denote position in sequence of PGHS-1 only) the corresponding peptide chosen for PGHS-2 will be addressed to selectively inhibit folding of PGHS-2. (i.e., Significant binding affinity of many bivalent agents will only occur for the protein in which both sites are available. Binding to only one site may occur with an affinity several orders of magnitude lower.)

Hence, as per diagram 1, sequences in PGHS-2 (the target sequences) corresponding to the above noted sequences of PGHS-1 are as follows.

1 PEHLRFA (compare this to the corresponding sequences in PGHS-1, PPQSQMA)

2 GDEQLFQ (this sequence is identical in both PGHS-1 and PGSH-2)

Diagram 1

Alignment of Sheep PGHS-1 (Query) and Human PGHS-2 (Sbjct) by NCBI Blastp, Based on Sequence Homology.

(Format is same as actual blast readout.)

```
Score =  734 bits (1875), Expect = 0.0
 Identities = 332/553 (60%), Positives = 422/553 (76%), Gaps =
1/553 (0%)

Query: 10  NPCCYYPCQHQGICVRFGLDRYQCDCTRTGYSGPNCTIPEIWTWLRTTLRPSPSFIHFLL 69
           NPCC +PCQ++G+C+  G D+Y+CDCTRTG+ G NC+ PE  T ++  L+P+P+ +H++L
Sbjct: 19  NPCCSHPCQNRGVCMSVGFDQYKCDCTRTGFYGENCSTPEFLTRIKLFLKPTPNTVHYIL 78

Query: 70  THGRWLWDFVN-ATFIRDTLMRLVLTVRSNLIPSPPTYNIAHDYISWESFSNVSYYTRIL 128
           TH +  W+ VN   F+R+ +M  VLT RS+LI SPPTYN  + Y SWE+FSN+SYYTR L
Sbjct: 79  THFKGFWNVVNNIPFLRNAIMSYVLTSRSHLIDSPPTYNADYGYKSWEAFSNLSYYTRAL 138

Query: 129 PSVPRDCPTPMGTKGKKQLPDAEXXXXXXXXXXXXIPDPQGTNLMXXXXXXXXXXXXXX 188
           P VP DCPTP+G KGKKQLPD+              IPDPQG+N+M
Sbjct: 139 PPVPDDCPTPLGVKGKKQLPDSNEIVEKLLLRRKFIPDPQGSNMMFAFFAQHFTHQFFKT 198

Query: 189 SGKMGPGFTKALGHGVDLGHIYGDNLERQYQLRLFKDGKLKYQMLNGEVYPPSVEEAPVL 248
             K GP FT  LGHGVDL HIYG+ L RQ +LRLFKDGK+KYQ+++GE+YPP+V++
Sbjct: 199 DHKRGPAFTNGLGHGVDLNHIYGETLARQRKLRLFKDGKMKYQIIDGEMYPPTVKDTQAE 258
```

```
Query: 249 MHYPRGIPPQSQMAVGQEVFGLLPGLMLYATIWLREHNRVCDLLKAEHPTWGDEQLFQTA 308
           M YP  +P    + AVGQEVFGL+PGLM+YATIWLREHNRVCD+LK EHP WGDEQLFQT+
Sbjct: 259 MIYPPQVPEHLRFAVGQEVFGLVPGLMMYATIWLREHNRVCDVLKQEHPEWGDEQLFQTS 318

Query: 309 RLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGAQFQYRNRIAMEFNQLYHWHPLMP 368
           RLILIGETIKIVIE+YVQ LSGY  +LKFDPELLF  QFQY+NRIA EFN LYHWHPL+P
Sbjct: 319 RLILIGETIKIVIEDYVQHLSGYHFKLKFDPELLFNKQFQYQNRIAAEFNTLYHWHPLLP 378

Query: 369 DSFRVGPQDYSYEQFLFNTSMLVDYGVEALVDAFSRQPAGRIGGGRNIDHHILHVAVDVI 428
           D+F++  Q Y+Y+QF++N S+L+++G+   V++F+RQ AGR+ GGRN+   + V+
Sbjct: 379 DTFQIHDQKYNYQQFIYNNSILLEHGITQFVESFTRQIAGRVAGGRNVPPAVQKVSQAST 438

Query: 429 KESRVLRLQPFNEYRKRFGMKPYTSFQELTGEKEMAAELEELYGDIDALEFYPGLLLEKC 488
           +SR ++ Q FNEYRKRF +KPY SF+ELTGEKEM+AELE LYGDIDA+E YP LL+EK
Sbjct: 439 DQSRQMKYQSFNEYRKRFMLKPYESFEELTGEKEMSAELEALYGDIDAVELYPALLVEKP 498

Query: 489 HPNSIFGESMIEMGAPFSLKGLLGNPICSPEYWKASTFGGEVGFNLVKTATLKKLVCLNT 548
           P++IFGE+M+E+GAPFSLKGL+GN ICSP YWK STFGGEVGF ++ TA+++ L+C N
Sbjct: 499 RPDAIFGETMVEVGAPFSLKGLMGNVICSPAYWKPSTFGGEVGFQIINTASIQSLICNNV 558

Query: 549 KTCPYVSFHVPDP 561
           K CP+ SF VPDP
Sbjct: 559 KGCPFTSFSVPDP 571
```

Probe Labeling ("probes" refer to embodiments in which screening of a plurality of test compounds, e.g., a combinatorial chemical library, is to take place.)

Peptide sequences (1 and 2 of those above) can be synthesized in an automated peptide synthesizer or manually utilizing standard solid phase peptide synthesis (SPPS) (see , e.g., Novabiochem Catalog and Peptide Synthesis Handbook, 1997/98, for protocols or Solid Phase Peptide Synthesis edited by Stewart (1969)) or purchased from a peptide synthesizing facility. The handling of peptides (e.g., dissolving, purification by gel filtration and HPLC, and storage) are described in the above reference and will be familiar to anyone skilled in the art of peptide synthesis (also see, Rivier, J. et al (1984) J. Chromatog. 288, 303-328 as general reference.)

The purified peptide is preferably labeled (for assay purposes), e.g., with dye, either (but not limited to) the chromogenic dyes, Disperse Red 1 or Disperse

Blue (Aldrich), or alternatively (but not limited to) the fluorescent dyes, fluorescein isothiocyanate (FITC) and tetramethylrhodamine-5 (or 6)-isothiocyanate (TRITC). For the chromogenic dyes, this is done by first acylating the dye with glutaric anhydride to yield dye-$CO(CH_2)_3COOH$. (see Rustum, B. et al. J. Am. Chem. Soc. 1994, 116, 7955-7956. This publication will henceforth be referred to as "**reference 1.**") Peptide labeling is then accomplished by N-acylating the resin-bound peptide with the acylated dye, to yield

dye-$CO(CH_2)_3CO$-(probe sequence 1 or 2).

Fluorescein and rhodamine labeled peptide is produced as per protocol (See, e.g., Bioconjugate Technologies by G.T. Hermanson (Academic Press 1996), chapter 8, "Tags and Probes," pp. 304-305, 320). However, if the peptide contains a lysine or arginine, the side chain amines must remain protected during labeling. So, for Fmoc peptide synthesis Lys(Dde) and Arg(Mtr) may be used. Here the dye may be added to the resin bound peptide (i.e., polyethylene glycol grafted polystyrene resin) and the above ligation protocol used.

The probe is used to perform assays (to be described) in either organic or aqueous solvent. This will depend on the solubility of the probe and on the swelling characteristics of the resin. If a probe is insoluble in water and it is desired to perform a binding assay in water, the probe can be modified to improve its solubility:

The original peptide probe sequence can, for example, be altered, by adding at its N- terminus (during standard SPPS): ser-gly-gly-arg-arg, yielding:

dye-$CO(CH_2)_3CO$-(probe sequence)-ser-gly-gly-arg-arg (3)

Here ser-gly-gly acts as a flexible linker (optional) separating the probe sequence from arg-arg which is used to increase water solubility. Thus, the peptide part of **3** is synthesized as in **1** and **2**, but the probe sequence contains the additional 5 N-terminal amino acids.

Labeling a longer polypeptide chain

Several methods can be used to label a protein or longer polypeptide chain. The protein may, for example, be labeled with the fluorescent dyes, FITC or TRITC, at primary amines, as per above protocols. See Bioconjugate Technologies by G.T.

Hermanson (Academic Press 1996), chapter 8, "Tags and Probes," pp. 304-305 for protocol and this is the preferred method in the present invention.

As an alternative, the protein may be radiolabeled. *E. coli* cells expressing the target protein or polypeptide in recombinant form as an insert on a selectable plasmid vector can be grown in culture with ^{14}C-labeled leucine, lysine, phenylalanine, tyrosine, or ^{35}S-methionine or cysteine to radioactively label the target protein (first it should be ascertained from the amino acid sequence of the protein that it contains the amino acid used for labeling (see, e.g., Methods for General and Molecular Bacteriology edited by Gerhardt, Murray, Wood, and Krieg. American Soc. for Microbiology (Washington, D.C.) 1994. p. 514, 22,1,2). The target protein is purified (see, e.g., Protein Purification Protocols, Doonan, editor. Humana Press (Totowa) 1996).

Alternatively the protein may be produced in a eukaryote cell and, depending on amino acid content, labeled with ^{35}S -methionine and/or cysteine, or as above. In some instances the protein or peptide may be iodinated with radioactive iodine, ^{125}I.

Before use in a bead binding (screening) assay (to be described) the labeled target protein (or polypeptide greater than 16 amino acids) is preferably dissolved in aqueous buffer (or water) containing 2-3% sodium dodecylsufphate (SDS, Sigma) and heated to 95°C for 5 minutes. (protein concentration should be kept below 50mg/ml) If the protein contains cysteine(s), 5mM betamercaptoethanol (Sigma) is included. The treated protein solution is diluted 20:1 in aqueous non-denaturing buffer containing 0.1-1.0% Triton X-100.

Combinatorial Chemistry

During the last decade enormous progress has been made in the field of organic chemical syntheses concerned with the generation and screening/analysis of chemical diversity. Collectively these technologies are referred to as combinatorial chemistry. The impetus for their development, apart from the desire of chemists to improve and expand synthetic methods, was the need within the pharmaceutical and biotechnology industries to accelerate drug discovery and development programs through the rapid generation and mass screening of enormous numbers of lead compounds, and more subtly to reduce the cost of drug development by increasing the likelihood that lead compounds produce viable drugs. To date combinatorial

chemistry has accelerated lead compound discovery by 1) rapid synthesis of very large libraries of compounds, 2) rapid, parallel screening, and 3) by facilitating structural elucidation of leads and providing means for their subsequent re-synthesis.

The first examples of combinatorial chemistry involved synthesis of peptides and nucleic acids. Peptide combinatorial libraries have been synthesized biologically using bacteriophage lambda, and chemically on solid supports based on the methods of Merrifield. Nucleic acids have also been produced chemically on solid supports. Subsequently techniques have been expanded to include synthesis of other polymeric compounds such as peptoids (N-alkylated glycines), oligocarbamates, oligoureas, vinylogous peptides, peptidosulfonamides, vinylogous sulfonyl peptides, azatides, and ketides. In general these syntheses involve the sequential incorporation of orthogonally protected bifunctional chemical building blocks containing sites of diversity that are serially protected and deprotected to either prevent or allow chemical modifications as dictated by each step of synthesis. These methods have been extended to allow chemical diversification of a core molecule or scaffolds such as biphenol, 3-amino-5-hydroxybenzoic acid, cyclohexane, acylpiperidine, cholic acid and many others. Diversification schemes can involve synthesis of oligomers, or non-oligomeric small molecules through the use of a varied set of bifunctional organic molecule building blocks.

One of the most common and efficacious methods of synthesizing combinatorial libraries involves the split synthesis method on a solid support, or the "one bead, one compound" set of methods. Here an initial component is derivatized, generally to polymer beads such as cross linked polystyrene, polyethylene glycol, acrylamide, or composite resin either by binding directly to a functionality of the bead or using a conduit chemical as a linker. Beads are divided into aliquots, each being reacted with a specific orthogonally protected chemical building block, re-mixed, re-divided into aliquots, deprotected and each new aliquot subsequently reacted with another orthogonally protected chemical building block; the cycle is repeated until completion of the desired library. As a result each bead contains a single library compound and libraries may contain enormous numbers of different compounds. Detachable linkers allow solution phase manipulation and assay of library compounds. On-bead binding assays are also used for identifying hits. The structural identities of library compounds is assessed by direct chemical analysis or

by incorporation of tags on each bead at each step of library synthesis. Such tags are easily analyzable and represent a synthetic history of each library compound. Other means of identifying library members employ spatial address systems.

One Bead-One Compound Methodology

The choice of polymer bead resin depends on several criteria. First of all the resin must be compatible with the chemical reactions and reactants used in the synthesis (including the coupling and decoupling of compounds to and from the resin), as well as with the tagging system used for compound identification. Uniformity of bead size and loading capacity are also important factors. Various resins are able to swell in either aqueous or organic solvents, or both. The swelling of a resin acts to solubilize the chemical building blocks that are attached to it, and therefore the choice of resin affects the methods of screening compounds in on-bead assays, as well as the range of usable chemical reactions. Screening in aqueous solvent for example will allow one to approach physiological solution conditions and thus the physiological behavior of a biological target (e.g., protein targets and cells are not appropriate for screening in organic solvent). TentaGel™ (polyethylene-grafted polystyrene) is an example of a resin that swells both in water and organic solvent.

The split synthesis method involves dividing the free resin (or resin conjugated to a linker) into i aliquots (where i equals the number of different building blocks initially coupled to the resin. i is equal to or greater than 1.) followed by addition of i building blocks (one per aliquot). The resulting reaction supplies the first set of protected chemical building blocks and also may serve to link the library compounds to the resin or to a scaffold molecule or linker that has already been linked to the resin. Subsequent synthetic cycles may involve varying numbers of building blocks and synthesis steps. Cleavable linkers allow release of library compounds which may facilitate various kinds of assays. (Cleavable linkers are not required where on-bead binding assays serve to identify initial leads and where there is no need to work with free compound for further analysis and assay or where such analyses may be deferred).

Synthesis of a library is usually followed by screening and decoding, which serve to 1) identify library compounds with a desired property (e.g., affinity for a

target) and 2) identify their chemical structures, respectively. Screening is accomplished either by binding a target directly to bead-associated compounds (on-bead assay) or by releasing compounds from beads and assaying in solution. The choice of tagging system (for decoding) will depend on its compatibility with the chemistry of the library and on its ease of analysis. Chemical tags involving binary coding (e.g., halobenzenes or secondary amines) are very useful. Essentially tags are incorporated onto the beads at each step in library synthesis. The set of tags associated with a given bead thereby records the synthetic history of the unique library compound bound to a given bead.

Combinatorial Libraries Yielding "Synthetic Receptors" that Bind Peptides with Specificity and Selectivity

Nature has had no problem creating macromolecules that bind peptides with high affinity and specificity. Anti-peptide antibodies are an example. Furthermore, even those antibodies that bind the surfaces of whole proteins have for the most part been shown to bind specifically to short, flexible peptide sequences exposed on the protein's surface. Additionally, many other natural receptors bind peptides; e.g., the substance P receptor, the tuftsin receptor, the neurotensin receptor and opiod receptors. Synthetic receptors that bind peptides have also been produced, through combinatorial chemistry. These include, for example, bucket shaped macrocyclic compounds and other compounds built around various rigid molecular scaffolds.

In particular, "molecular tweezers" as produced by Still and others (Rustum, B. et. al. J. Am. Chem. Soc. 1994, 116, 7955-7956) consist of a cholic acid scaffold appended by combinatorially diversified tripeptides, a structure which combines a design motif (rigid scaffold) with chemical diversity. These compounds exhibit micromolar affinities for various tetra- and penta-peptides. Furthermore, Nestler showed that conformational rigidity of the scaffold was necessary for selectivity and affinity of binding (Nestler, H.P. Molecular Diversity, 1996, 2, 35-40).

A variety of other "two-armed" and bucket shaped receptors have been produced utilizing other types of rigid chemical building blocks. Additionally, several investigators have shown that various organic dyes bind selectively to specific peptides: this can be interpreted to mean that rigid organic molecules lacking predetermined or rational design constraints, such as bucket shape or

tweezer-like design, have the ability to bind flexible peptides with varying degrees of selectivity. The common structural requirements are a degree of conformational rigidity and complexity of structure. However, in all of these examples, the highest affinity binding is in the micromolar range. This is relatively low when considering a molecule for candidacy as a drug but is consistent with the flexibility of peptides which imposes a high entropic price for binding.

Synthesis of Combinatorial Library Based on a Cholic Acid Scaffold

Combinatorial library synthesis using the "one bead-one-compound" approach is conducted with laboratory hardware familiar to those skilled in the art (See for example Multiple Peptide and Non-Peptide Libraries, chapter 4. G. Jung, editor. VCH 1996).

Cheno(12-deoxy)-cholic acid is bound to (aminomethyl)-polystyrene beads (Novabiochem's AM Resin) using diisoproplycarbodiimide (DIC) as in standard peptide bond formation (see **reference** 1) or alternatively PyBOP (Novabiochem), TBTU, or HBTU may be used as per protocol (see Novabiochem Catalog and Peptide Synthesis Handbook, 1997/98, Synthesis notes). Approximately 3 x 10^6 beads are used for a library that will consist of 1 million compounds. Swelling of the resin is done in dimethylformamide (DMF) and a 4 x molar excess of cholic acid is added (based on resin capacity; see Novabiochem, The Combinatorial Chemistry Catalog; and Catalog and Peptide Synthesis Handbook, 1997/98). The product of the above reaction constitutes a steroidal scaffolding bound to the resin through an amide bond, wherein said scaffold yields two sites of diversity at the C(3) and C(7) hydroxyls of the steroid, respectively. These two sites of diversity will be appended, each by a combinatorially derived peptidic arm similar to that described in **reference** 1.

The cholic acid may otherwise be added to NovaSyn TG amino resin (amino-functionalized polyethylene glycol-grafted polystyrene beads) where it is desired to assay for probe binding under aqueous conditions. An applicable coupling reaction protocol is described in Methods in Enzymology volume 267, 1996. Chen, C.L et al. pp. 211-247, herein called **reference** 2. Alternatively, the coupling can be done in DMF as above and synthesis of the combinatorial library

can proceed as with polystyrene resin. After synthesis on TG or other water swellable resin, the library may be screened in aqueous buffer (to be described).

Combinatorial synthesis proceeds as follows: The C3 hydroxyl of the cholic acid scaffold is selectively acylated by one of each of the following protected Fmoc-(L) amino acid fluorides: alanine, valine, leucine, phenylalanine, proline, serine(tBu), threonine(tBu), lysine-(Boc), aspartic acid(tBu), and glutamic acid(tBu), where addition of this first amino acid is treated as the first step in three rounds of combinatorial synthesis wherein the same Fmoc-amino acids are used as randomized building blocks in each subsequent round of synthesis. The Fmoc-amino acids fluorides are produced by reaction of the protected amino acid with cyanuryl fluoride as per Bertho et al. Tetrahedron Letters, 32,10, 1303-1306, a.k.a. **reference** 5. Acylation at the C3 site is carried out by initially dividing the cholic acid conjugated resin into 10 vessels and adding 1.5 equivalents of the respective Fmoc-amino acid fluoride in dimethylformamide (DMF) (see **reference** 1). Two additional rounds of combinatorial synthesis as stated above are carried out according to **reference** 1 and 2. The last amino acid is capped with AcOH as in **reference** 1 (see Solid Phase Peptide Synthesis edited by Stewart (1969) p. 33 using acetic anhydride) yielding compounds with one peptidic arm:

P-NH-scaffold-C3(aa_1-aa_2-aa_3-Ac, V_2=H), where P= bead, Ac=acetyl, V= second diversity site of scaffold, aa=amino acid, H=hydrogen.

Encodement of the library is accomplished by using either the binary tagging system as per Ohlmeyer, M.H.J, et al. PNAS 90, 10922-10926 (1993). a.k.a. **reference** 3 or that of Borchardt, A., Still, W.C. J. Am. Chem. Soc. 116,373-374. (a.k.a. **reference** 4 [whichever is more convenient]). Other tagging systems may also be used (Ni, Z.J. et al. (1996) 267, 261-272, a.k.a. **reference** 6). The number of tags used for each peptidic arm (in any case) will be 12. So, a total of 24 tags (synthesis of tags according to above references) will be used. Later decoding of selected beads will involve liberation of tags and analysis as per **references** 3, 4, or 6 as per above.

The second combinatorial synthesis (for construction of the second peptidic arm is begun, first by acylating the C7 of the cholic acid scaffold by using 4-(N,N diethylamino)pyridine as a catalyst (see Caprino, L.A. et al. J. Org. Chem. 1991, 56,

2611, 2635 and **reference** 1) and utilizing, as before, each of the ten noted Fmoc-(L) amino acid fluorides. This is again followed by two additional rounds of combinatorial synthesis as above (as per **references** 1 and 2) followed again by capping (N-acetylation) of the last amino acid. Here the second set of 12 tags are employed as per the previous tagging. Before screening the library compounds, herein called "receptors," amino acid side-chain protecting groups are now removed with 25-95% trifluoroacetic acid (TFA) as per Novabiochem Catalog & Peptide Synthesis Handbook 19997/98

Screening the Library

If the library is synthesized on polystyrene beads, it is screened in organic solvent (e.g., $CHCl_3$) as per **reference** 1. If the library has been synthesized on polyethylene glycol/polystyrene composite beads (e.g., Tentagel or Novabiochem amino TG resin) or some other water compatible bead resin, the assay may be done either in organic solvent or in water (if in water use washing and blocking protocol in **reference** 2 (refer to pp. 217) however, incubate the probe with the beads (for labeling) in an aqueous buffer such as 10-50mM Tris pH 7.4, with 100-150mM KCl or NaCl (or the acetates), or other incubation media of similar ionic strength and buffering).

Additionally, one mode of aqueous screening is carried out as above but with 0.1-1.0% Triton-X 100, and 1mM DTT or 5mM beta mercaptoethanol (one of the latter two ingredients if probe contains cysteine). This mode of screening should be used when the probe consists of an entire polypeptide chain or peptide longer than 16 amino acids (as noted earlier), which has been unfolded (prior to screening) by heating to 95°C in 2-3% sodium dodecylsulfate (SDS) for at least 5 minutes. This is followed by no less than a 20:1 dilution of the SDS dissolved probe in an aqueous buffer containing 0.1-1.0% Triton-X 100, as noted earlier. It may also be used for shorter peptides if desired as a method of enhancing probe solubility in water (and in place of synthesizing a probe with ser-gly-gly-arg-arg, as previously described).

Screening of the library is done individually for each probe (there will generally be two) by incubation as described in **reference** 1 and/or **reference** 2 (except for probe concentration --see below). Thus, at least two "copies" of the library will have to be synthesized. This can be done, for example, by using two

"tea bags" (see Multiple Peptide and Non-Peptide Libraries, G. Jung, editor. pp. 82-83. VCH 1996) in each reaction vessel (during library synthesis) so that only one synthesis cycle need be employed. The concentration of probe during incubation is preferably set at 1μM. Intensely colored beads can be detected by visual inspection and manual isolation performed (**reference** 2). If the probe contained fluorescein or other fluorophore, detection can be done under UV illumination (Mossberg, Ericsson previous reference). If no colored (stained) beads result, incubation with labeled probe should be repeated using higher concentrations of probe (e.g., 10μM and if necessary, 100μM).

Intensely colored (stained) beads (between 1 and 100 beads; or number at the discretion of the assayer) are picked for decoding as previously described.

If probe contains radioactive label, use detection protocol as per **reference** 2 pp. 214-215 (employing longer exposure time based on radionuclide).

Ascertaining Selectivity

In general the most selective receptor will be the most useful. The control of selectivity involves both selection of probes as previously discussed and the screening assay itself. For example, if the drug developer or assayer wishes to have a receptor that distinguishes between a peptide corresponding to oncogenic H-ras (val12) and one corresponding to normal H-ras (gly12), the original library would be screened with two probes representing homologous sequences from the proteins one wishes to distinguish (e.g., probes containing peptide 1, amino acids 6-17 of oncogenic and normal H-ras proteins). One probe is labeled with Disperse Red as described; the other is labeled similarly with Disperse Blue and a two color binding assay is employed as in **reference** 1. Here, a doubly stained bead (purple) would mean that the receptor on that bead was not able to distinguish between the two peptides (signaled by it binding to both probes). However, beads colored either intensely red or blue would contain receptors that are selective for one or the other of the two peptides; i.e., a bead stained only, or mainly, by the probe for oncogenic ras (e.g., red or reddish purple) would represent a potentially useful receptor because it would be selective for oncogenic ras (if this distinction was desired). Double staining using FITC and TRITC can be employed similarly (for detection and

interpretation of double signal see Mossberg, K., Ericsson, M. (1990) Journal of Microscopy. 158 (Pt 2):215-24).

A means of increasing the likelihood that a given set of receptor hits will contain a higher proportion of receptors that distinguish between two or more close variations of a given peptide (as in the ras example) is to use shorter probe sequences for library screening. For example, in the case of H-ras, a probe consisting of the shorter sequence, AVGVGK, could be used in place of probe 1. The single amino acid substitution separating normal H-ras (gly 12) from oncogenic H-ras (val 12) would then comprise a larger part of the probe sequence. Shorter probes contain on average fewer potential binding sites for receptors (library compounds); thus, in a shorter probe, the distinguishing mutation (or variation) will be present in a larger proportion of the (fewer) available binding sites. This increases the probability that a given hit is selective. At the same time, longer probe sequences will increase the number of positive beads (by affording more potential binding sites). This results in hits by a greater number of selective and non-selective receptors; thus more time may be involved for screening (to find a selective receptor) but there is also a higher likelihood of finding a selective, high affinity receptor.

If a probe contains additional components (e.g., extra amino acids) not part of the target sequence, it will be necessary to exclude the possibility that binding was due to these extra components. This is tested, for example, by decolorizing the bead in DMF (to remove bound probe which is then washed away with the liquid phase; see Lam, K.S., Lebl, M., Krchnak, V.Chemical Reviews 97(2) 347-510); and then incubating the decolorized bead with the extra amino acid component linked to the original dye (e.g., dye-ser-gly-gly-arg-arg, synthesized by the same methods noted for labeled probes). If the bead is re-colorized, binding may not be specific for the target sequence, and a different receptor must be chosen for testing from the available hits (or from a re-screening). To assure a greater possibility of selectivity, recolorization can be tested by using the original probe peptide linked to a different dye with a different solubilizing sequence (e.g., gly-gly-glu-glu) Alternatively, use a dual color system as per Lam. K.S. et al. J. Immunol. Meth. 180 (1995) 219-223.

Selectivity also reflects the ability of a given receptor (that has been shown to bind to a given peptide probe) to bind other peptides whether they are present or

not in the target or related proteins. In other words, if the receptor binds a given tetratapeptide, how many different pentapeptides or other peptides will it bind? This aspect of selectivity is ascertained by screening the free receptor against a combinatorial peptide library (e.g., pentapeptide library built using the 20 proteogenic amino acids) wherein the peptide library is synthesized on beads according to the method of **reference** 2, with the following modification. No matter how long the probe, the receptor is not likely to bind more than 5 amino acids. So, a pentapeptide library is a sufficient approximation offering a great diversity of potential pharmacophores. Additionally, the library need not be complete (because of redundancy in such libraries). Therefore it will be sufficient to synthesize the pentapeptide library using 10^7 beads (see Multiple Peptide and Non-Peptide Libraries, G. Jung, editor. pp. 82-83. VCH 1996, for details).

Here the receptor will be used to screen the combinatorial peptide library. The receptor structure and its synthesis history will be known because this was previously decoded from the earlier bead tags. The receptor is then resynthesized and labeled on Tentagel TGR resin which will allow cleavage of the receptor from the resin by acid (see Novabiochem Catalog and below and take precautions in case released compound becomes insoluble; e.g., redesolve in organic solvent), followed by purification which involves the filtration, HPLC and lyophilization techniques commonly used in peptide purification (see below). However, before addition of the cholic acid scaffold, the resin is derivatized with Fmoc-Lys(Dde)OH followed by Fmoc deprotection and addition of the cholic acid to the lysine primary amine.

[This receptor can also be synthesized on sulfamylbutyryl resin (to be described later). In this case, removal of the receptor from the resin involves exposure to hydroxide. Under these circumstances various receptors will be soluble. Therefore this means of synthesis may be preferable in some instances where solubility problems occur.]

If the receptor itself contains lysine or arginine, its side chain protecting group should be Mtt (e.g., Fmoc-Lys(Mtt)OH) and this should not be removed until after labeling of the receptor with the dye. Labeling of the receptor is done with FITC or TRITC, using the previously noted protocol, however, it is done while the receptor is bound to the resin and it will occur on the lysine side chain amine which is first deprotected with 2% hydrazine as per Novabiochem protocol (see earlier

mentioned reference). Mtt side chains (if present) can then be removed with TFA as per Novabiochem protocol (actually at the same time receptor is cleaved from resin by acid) and the labeled receptor removed from the resin as per above reference, filtered and purified (see Peptide Purification, Novabiochem Catalog and Peptide Synthesis Handbook p S58); note, the receptor will generally require organic solvent for dissolving. (This "receptor" is not yet the "lead compound," which will contain two "receptors" where a bivalent format is used.)

Then the labeled receptor is screened against the peptide library as described in (Torneiro, M., Still, W.C. J. Am. Soc. 1995, 117, 5885). The most selective receptors will be those binding to the fewest peptides in the peptide combinatorial library. Thus, the choice of receptor for further development will be based on a relative measure of selectivity. Very high selectivity at this point is not necessary for development of bivalent or multivalent compounds, though high affinity receptors having the most selectivity should preferably be chosen. Selectivity will be determined more critically by the bivalency of the lead compound (below) and a lead compound consisting of two moderately selective receptor valencies can still exhibit high selectivity, as a bivalent unit.

An alternative cholic acid scaffold library can be constructed as per **reference** 1 where the first amino acid at the C3 and C7 sites is glycine and where the peptidic arms contain four amino acids each. A further embodiment consists of an A,B-trans-steroidal core as in Cheng, Y., Suenaga, T., Still, W.C., (1996) J. Am. Chem. Soc. 118,1813-1814, where the present invention differs only in having tripeptidic, in place of dipeptidic, arms. This library produces remarkably selective receptors due to their conformational rigidity.

A Bivalent Lead Compound

For a given target protein, two receptors from the previous cholic acid library are chosen. These either bind to different peptide sequences within the target protein or they may bind to the same sequence which is repeated in the target polypeptide. In this latter case the two receptors can be identical and only one screening would have been required. The assayer will be presented with a set of potential receptors to choose from for subsequent synthesis of the bivalent lead compound. The choice will be based on the affinity and selectivity of a receptor for

its peptide target. It is sufficient to estimate relative affinity (non-quantitatively) by observing the relative intensities of colored beads (from the previous screening). The most intensely colored beads will contain receptors of highest affinity binding. Selectivity is ascertained in the ways previously described. The receptors chosen should have relatively high affinity and exhibit selectivity, as well. Ideally, a high affinity receptor having the greatest selectivity should be chosen. In practice the most selective receptors do not always have the highest affinity binding.

To make the bivalent receptor, which constitutes the lead compound, the two previously defined receptors (call them arbitrarily receptor 1 and receptor 2) must be linked without disturbing the peptide binding pharmacophore of each. This is done, for example, by resynthesizing receptor 1 on Novabiochem's 4-Sulfamylbutyryl AM resin. Attachment of the cholic acid scaffold to the resin is activated by diisopropylethylamine (DIPEA) in DMF as per Novabiochem's document, "Suggested Technical Tip #2" (coupling to resin). Remaining Fmoc synthesis and side chain protection with acid labile protecting groups is done, as before. After synthesizing the receptor according to the previous decoded synthesis, it is detached from the linker with hydroxide (see protocol: Brakes, B.,J., Ellman, J.,A. (1994) J. Am. Chem. Soc. 116, 11171-11172 and Brakes, B., J., Virgilio, A.,A., Ellman, J.,A. (1996) J. Am. Chem. Soc. 118, 3055-3056) (but amino acid side chain protecting groups are left intact) leaving an unprotected COOH moiety. The receptor is then filtered and purified as previously noted, and is lyophilized and dissolved in DMF.

Receptor 2 is synthesized on Novabiochem TGR (amide resin with detachable linker) with the following modification. Before the steroid is added to the resin, the resin is acylated with Fmoc-lysine(Dde)OH. Then after Fmoc deprotection (piperidine in DMF as before) the cholic acid scaffold is added to the N^{α}-NH_2 functionality of the lysine residue by activation with DIC, or other common coupling reagents, as previously done for addition to aminomethyl derivatized resin. The remaining portion of receptor 2 is then synthesized as before (with acid labile side chain protecting groups) and left protected except for terminal amino acids which are N^{α} deprotected and capped as before.

Then the Dde protecting group of the lysine side chain is detached with 2% hydrazine (see Novabiochem Handbook for details). Free COOH-receptor 1 (previously synthesized on sulfonamyl resin (above)) is added in DMF and an amide

bond between receptor 1 and the unprotected lysine side chain NH_2 is formed using DIC, or other common coupling reagent to activate (as per SPPS). The bound compound (Receptor 1-Receptor 2) which is now a bivalent receptor dimer, is removed from the resin by treatment with TFA (Novabiochem Catalog as before) and at the same time the TFA treatment removes all of the remaining (acid labile) side chain protecting groups. The bivalent lead compound is then purified as previously noted for purification of peptides (e.g., purification using reverse phase HPLC and lyophilized). The dried compound is then dissolved in either water, aqueous buffer, methanol, ethanol or dimethylsulphoxide (DMSO) and set aside as an analyte for biological testing as a lead compound.

A generalized structure of such a peptidosteroidal bivalent compound is shown in Figure 2.

The dimeric steroid is likely to be soluble in water (see Davis, A. P., Chemical Society Reviews 1993, 243-253.) and due to its lipophilic steroidal components is also likely to enter cells. Hence, this structure in addition to fulfilling the folding inhibitory requirement of a lead compound (assays to be described) is also likely to posses favorable bioavailability properties.

RECEPTORS DERIVED FROM A HETEROGENEOUS LIBRARY

Cholic acid is bound to 8.5×10^6 NovaSyn TG amino resin beads as previously, and is divided into 8 parts. A library consisting of approximately 2.8×10^6 compounds will be produced. The C3 OH of Cholic acid is acylated (as previously) with one of each of the following 8 Fmoc-amino acid fluorides (the fluorides can be prepared as noted previously): Gly, Ala, Val, Leu, Ile, Met, Phe, Pro. One of the previous binary coding systems is used, utilizing 24 tags as per **reference** 3. The resin is then remixed and divided into 6 vessels. Then C7 is acylated (as before) but with the following Boc-amino acids (and before any further synthesis at the C3 arm): Gly, Ala, Val, Leu, Ile, Met, Phe, Pro. Next, diversification of the C3 arm proceeds by Fmoc-amino acid deprotection (as previous); the resin is then remixed and divided into 10 vessels as per Krchnak, V. et al. Molecular Diversity, 1, (1995) 177-182. The protocol therein is followed for randomization and addition of 10 aromatic hydroxy acids followed by 21 alcohols

(side group protection not necessary). Next Boc deprotection is done (see Novabiochem Cominatorial Chemistry Handbook and Catalog & Peptide Synthesis Handbook), and the C7 amino acid is appended by combinatorial synthesis as above with the 10 aromatic hydroxy acids and 21 alcohols. Screening is done as per previous receptor library.

A bivalent receptor is produced as above by binding Fmoc-Lys(Dde)OH to Novabiochem TGR resin. The receptor (e.g., Receptor 2) which has been pre-synthesized with a COOH functionality, as previously on sulfonamyl resin is added to the lysine side chain amine after synthesis of receptor 1 and Dde removal as above. The bivalent receptor is detached from the resin by 95% TFA as previous and purified according to Krchnak, V. et al. (1995) Molecular Diversity 1, 149-164.

Heterogeneous libraries capable of producing peptide binding compounds also include libraries having: 1) other scaffolds (see Krchnak, V. et al. Molecular Diversity, 1, (1995) 177-182.), 2) scaffolds with more than one site of diversity, 3) more than one such scaffold at various steps in the synthesis, 4) the scaffold placed in any position --e.g., to create a branch in the middle of the compound, 5) no branching, based on differentially protected or reactive sites, and 6) a single oligomeric building block (defined as a homogenous library).

Peptide Combinatorial Libraries and Libraries of Cyclic Peptides

The screening of a combinatorial peptide library consisting of linear or cyclic peptides will yield peptides that bind peptides in target proteins. These are derived by screening with peptide probes, or denatured polypeptide as previously described. A combinatorial pentapeptide or hexapeptide library is synthesized according to the method of **reference** 2 (also see reference cited below), on TentaGel or polystyrene resin as noted earlier, using only 10^7 beads (for the hexapeptide library --for an incomplete but sufficiently representative library. To optimize "hits" where an incomplete library is used, sequential screening can later be carried out. Refer to Jung, G. editor. Combinatorial Peptide and Nonpeptide Libraries. A Handbook VCH (Weinheim 1996) p. 184). A hexapeptide library using all 20 proteogenic amino acids can be encoded by 30 tags (of the kind described) using a binary coding system. For a pentapeptide library 25 tags are required. One form of this random peptide library approach will utilize the D-amino acid isomers. Thus fewer building

blocks (less than 20) may be used if various D-isomers are unavailable. Mixtures of L and D amino acids may also be used in library synthesis.

Screening of the library using dye-labeled probes, as previously described, can be done in either aqueous buffer or organic solvent and criteria for selecting positives are as before.

Additionally, a "receptor" peptide thus identified can easily be resynthesized (SPPS) and labeled as were the original probes, to scan a second peptide combinatorial library (as described earlier) to ascertain selectivity. The construction of bivalent peptide receptors (to generate the lead compound), where this is done, is done according to the methods and formulations that are noted for complementary peptides in the next section.

A cyclic peptide library (hexapeptides preferred) see Multiple Peptide and Non-Peptide Libraries, G. Jung, editor. VCH 1996, pp. 222-223, 327-347, is also a method for the present invention. Choice of building blocks are as above, but start with an aspartic or glutamic acid (i.e., for NH_2 derivatized resins) or other side chain appropriate amino acid to allow for bead attachment. Screening is done as before (after cyclization) with previous probes in resin-compatible buffer. The syntheses outlined in the above source involve end-to-end cyclization where peptides are attached to resin through a side chain. Two cyclic peptide receptors can be linked to form a bivalent lead where both the resin bound cyclic peptide (receptor 1) and a free (detached from resin) cyclic peptide (receptor 2) each contain an additional but protected side chain, where one (after deprotection) is an amine (e.g., Lys) and the other is a carboxylic acid (e.g., Asp). Deprotection and amide bond formation as in the examples involving synthesis of bivalent receptors is carried out as previously noted. However, a single cyclic peptide can also be tested as a lead compound in subsequent assays to test folding inhibition.

Other methods of cyclization may be employed (see, e.g., Darlak, K. et al. (1994) In Peptides: Chemistry, Structure and Biology. Hodges, R. and Smith, J. (eds.) Leiden: ESCOM, pp. 981-983), and Lyttle, M.H. et al. pp. 1009-1011).

As recognized by those skilled in the art, other chemistries can be used for identifying on developing protein folding inhibitors, including other backbones for multi-valent compounds.

FOLDING INHIBITORS BASED ON COMPLEMENTARY PEPTIDES

It has been shown that short flexible peptides are able to bind to other peptides and various proteins with high affinity and exquisite selectivity. This is determined by the complementary hydropathy that relates various amino acid side chains and has also been applied to antigenic mimicry and protein engineering. Specifically, two peptides composed of sequences of amino acids having opposite, or complementary hydropathies will, in many instances, bind specifically to one another. Thus in many instances, one can generate "peptide-binding peptides" rationally. Hydropathies appear to be strongly correlated to the genetic code. Although this peculiar property of the genetic code has not been fully explained, it does appear that hydropathically complementary peptides can be predicted by reading, in reverse orientation, antisense codon sequences and specifying the newly encoded amino acids (Jarpe, M.,A., Blalock, J.,E. (1994) in Peptides: Design, Synthesis, and Biological Activity Basava, C., Anantharamaiah, G.,M. editors. Birkhayser (Boston)). Furthermore, synthetic complementary peptides may be composed of L- or D-amino acids, which provides an opportunity to control the stability of such an agent in a physiological environment.

If one begins with a peptide, e.g., VFWKAR, its codons (specified by mRNA) can be assigned as: 5'-guc uuu ugg aaa gcu cga-3'. The anti-sense strand is: 3'-cag aaa acc uuu cga gcu-5'; read in the 5'-3' direction (ucg agc uuu cca aaa gac) this strand codes for a second peptide: SSFPKD, which is predicted to specifically bind to the first sequence, VFWKAR. If one of these sequences was part of a larger (folded) protein sequence, then a peptide consisting of the complementary sequence could still specifically bind it provided that the complementary sequence was accessible (e.g., if it was a free C- or N-terminal peptide or a surface-exposed loop). Partial accessibility may result in low affinity binding. In folded proteins, not only are internal peptides generally not accessible, but many surface exposed peptides are only partially accessible for binding. Where complementary peptides bind a protein target with low affinity, avidity has been increased by producing a binding agent (receptor) consisting of multiple copies of the complementary peptide. This usually takes the form of a very large branched peptide (however, the target sequence is generally present only once per given protein).

Other methods of generating complementary peptides have been developed, and are based on known or measured physical properties (hydropathy) of amino acids. However, the use of peptide combinatorial libraries overlaps these more specific examples of generating peptide interactions because the members of large peptide combinatorial libraries will include such peptides as can be designed or designated as "complementary" by other methods.

Folding inhibitor lead compounds can be designed by producing hydropathically complementary peptides to sequences within a target protein and using the resulting peptides to bind unfolded target proteins during their biosynthesis. Consider the previous example of human oncogenic H-ras. Two sequences were chosen as target sequences: Sequence 1 consists of amino acids 5 to 16 and sequence 2 consists of amino acids 73 to 86: (The protein contains a single folded domain.)

1 KLVVVGAVGVGK (contains distinguishing mutation)

2 RTGEGFLCVAINN

In the present example the target sequences will each be reduced to 5 amino acids in length, designated 1' and 2' (N- to C-termini), though the longer sequences would also be sufficient (in general, smaller lead molecules, other factors being equal, are preferred over larger molecules):

1' GAVGV (amino acids 10 to 14)

2' RTGEG (amino acids 73 to 77)

Next, a complementary peptide is derived for each of the above. This is done by 1) assigning a 5'-3' mRNA coding sequence for each peptide (based on the genetic code) or by assigning the sequence based on the actual cDNA. Next the anti-sense mRNA strand 3'-5' is written and then reversed so that it reads 5' to 3' (written "-RNA"). Then the code of -RNA is translated to reveal a "complementary" peptide (or set of potential complementary peptides). In the following example, one complementary peptide has been chosen for each target peptide:

coding mRNA (for peptide 1'): 5'-ggu gcu guu ggu guu-3'

antisense RNA: 3'-cca cga caa cca caa-5'

-RNA (encodes peptide 1"): 5'-aac acc aac agc acc-3'

By the same method the -RNA for peptide 2' is:

5'- acc uuc acc ggu acg -3'

The encoded peptides are thus:

complementary peptide 1": (NH_2)-NTNST-(COOH)

complementary peptide 2": (NH_2)-TFTGT-(COOH)

These peptides are predicted to bind specifically to peptides 1' and 2', respectively.

Note that the genetic code is degenerate; each amino acid is encoded by more than one codon. Sometimes reading the reversed (5'-3') complementary nucleotide sequence (to derive a complementary peptide) will yield a stop codon. When this occurs, choose an alternative codon for the corresponding site in the coding sequence (coding mRNA) of the target peptide so that the stop codon (in -RNA) will now be substituted for by an amino acid encoding triplet (e.g., If a coding mRNA contained uua for leucine, resulting in an anti-sense sequence of aau; the -RNA would be the stop codon, uaa. To alleviate this, reassign the coding sequence for the original leucine to, for example, uug. Then the anti-sense RNA will be aac and the -RNA codon will be caa, which codes for glutamine). The degeneracy of the code also allows one to specify many unique coding sequences for a given peptide (even though these may differ from the natural nucleotide sequence encoding the target protein). Although all of these sequences will encode the same target peptide, they may lead to different complementary peptides, all of which are potentially valuable leads. Lead compound optimization can thus proceed by comparing the binding and/or folding inhibitory activities (assays to be described) of the various complementary peptides and their formulations.

An alternative method of defining complementary peptides is to use a computer program as per Fassima, G. et al. (1992) *Archives Biochem Biophys.* 296, 137-143. That reference also contains a table of target amino acids and their hydropathically complementary amino acids and can be used for design of complementary peptides.

Additionally, proteins themselves have been shown to possess complementary sets of peptides. Thus, complementary peptides can also be derived by analysis of self-complementary regions of mRNA (see Draper, K.G., (1989) Biochem. and Biophys. Res. Comm. 163, 446-470).

Complementary peptides derived in the specific ras example above are predicted to bind specifically to polypeptides containing peptides 1' and 2',

respectively. This is true whether the complementary peptides are all L-isomers or all D-isomers (or may be mixtures of the two). D-isomer containing peptides are preferred for testing and administration in vivo because they are more resistant to proteolysis than L-amino acid peptides. However, in the *in vitro* translation folding assays to be described, L-amino acid peptides are suitable to test for selective high affinity binding and/or inhibition of folding without danger of degradation of peptides and this will be convenient if it is easier for the chemist or assayer to purchase a greater variety of protected L-amino acids (suitable for SPPS). D-amino acids should be used for assay *in vivo* or in cells or in systems wherein the ubiquitin or other proteolytic systems are reconstituted or otherwise active. The addition of protease to the folding assays to be discussed (limited proteolysis) does not, however, influence the choice of L or D-amino acids. If folding inhibition is shown to occur for a given L-amino acid containing peptide, the corresponding D-amino acid isomer is expected behave similarly.

Solid Phase Synthesis of Complementary Peptides and Their Formulations as Lead Compounds and Drugs

In one embodiment, a peptide complementary to at least one contiguous sequence within the target protein of, but not limited to, 4 to 16 amino acids is synthesized. In another embodiment, two peptides of 4 to 16 amino acids in length (5 to 6 #aa are preferred in the present invention) are chosen based on the above design method, wherein peptides are complementary to at least two target (peptides) sites within a target protein as previously described. A fusion or bivalent peptide is then synthesized using automated or manual solid phase peptide synthesis (SPPS), formulated, for example, in any of the following ways.

Branched Peptide

Fmoc-L-lys(Dde)OH is attached to Novabiochem TGR resin as per Novabiochem Catalog & Peptide Synthesis Handbook. A complementary peptide (e.g., complementary peptide 1") is synthesized (addition to first lysine) according to standard Fmoc peptide synthesis using tBu and Boc side chain protecting groups, however, the last peptide is Nα-Boc protected. Then the lysine (amino acid bound to solid support) is side chain deprotected with 2% hydrazine as per above reference.

Next, the second peptide (e.g., complementary peptide 2") is synthesized as a branch of the already synthesized peptide by attaching the first amino acid of 2' to the lysine side chain amine (now deprotected) by standard SPPS, using Fmoc protected L-amino acids with tBu and Boc protected sided chains. The last amino may be Nα–Boc or Fmoc protected. If it is Fmoc protected, release of Fmoc with piperidine (as previously) is required before the next step. Otherwise deprotect all acid labile protecting groups and simultaneously detach from resin using 95% TFA. Purification of the branched peptide can be carried out as per previously noted guidelines and protocols for peptide purification, and the purified peptide can be dissolved (for use as an analyte) in either water, aqueous buffer, DMSO, methanol, or ethanol as previously described for other lead compounds. The structure of the resulting peptide, which constitutes the lead compound is shown in generalized form in Figure 3. (The choice of which peptide is synthesized as the side chain branch is unimportant.)

Other standard side-chain attachments can also be used (e.g., where the first amino acid contains an acidic side chain).

A generalized structure of a bivalent branched peptide protein folding inhibitor is shown in Figure 3.

Fusion Peptides

The previous peptide may also be synthesized as a single fusion peptide by SPPS. In this case, the peptide order (which sequence comes first) is unimportant. A spacer may be synthesized separating the first and second sequences. This may consist of three glycines. The structure of a generalized bivalent fusion peptide is shown in Figure 4.

End-to-End Cyclic Fusion Peptides

A peptide containing a desired sequence (for example, as sequence as described above; see also Figure 5) is synthesized on Novabiochem Oxime resin using Boc-amino acids with acid insensitive side chain protecting groups (e.g. Fmoc). An additional three glycine residues (N-terminal) are added to the peptide following peptide 2' (by SPPS). End to end cyclization and concomitant release of the peptide from the resin is accomplished as per Osapay, G. et al. (1990)

Tetrahedron Letters, 31, 6121. Purification is done as per above reference and deprotection of side chains (which may differ from peptide of reference) are done as per previous protocols cited, after the cyclization/release and initial purification; final purification (where side chain deprotection is required) and solubilization of peptide for analyte, done as noted previously.

A structure of a generalized cyclic fusion peptide protein folding inhibitor is shown in Figure 5.

Alternatively, a circularized peptide containing two additional cysteine residues (preferably at or near the ends of a linear peptide) can be made by oxidation to cysteine or (without addition of cysteines) through amide bond formation by end-to-side chain ligation, by modifying the resin to contain an aspartic (or glutamic) acid linker (which would be added to the original sequence) (See, for these and other methods, G. Jung, editor. Multiple Peptide and Non-Peptide Libraries VCH 1996, pp. 222-223, 327-347, for examples of both.). Glycine spacers may minimize the possibility of producing a "neo-sequence" at the fusion junction; however, circular or other fusion peptides might be made with non-peptide aliphatic, or other, linkers. Similarly, a neo-sequence might be avoided in a branched, bi-valent peptide. Alternatively, a spacer may not be necessary and may be omitted.

Cyclic Peptides Complementary to a Single Continuous Target Sequence

Complementary peptides targeting continuous peptide sequences such as loops within a target protein are also potential folding inhibitors. For example, a peptide loop located mainly within the core of the target protein is identified from a published account of the target protein's structure. (Loops connect secondary structural elements.) As an example, take the sequence KLVVVGAVGVGK which represents a loop peptide in H-ras. Next assign the complementary sequence by using the genetic code as previously.

coding mRNA: 5'-aag cuu guu guu guu ggu gcc guu ggu guu ggu aag-3'
anti-sense RNA: 3'-uuc gaa caa caa caa cca cgg caa cca caa cca uuc-5'
-RNA: 5'-cuu acc aac acc aac ggc acc aac aac aac aag cuu-3'
Thus, the complementary peptide is:
LTNTNGTNNNKL

This peptide is synthesized and circularized as per previous example. In the example below, synthesis occurs on oxime resin, followed by release/circularization and purification, as above to yield the cyclic peptide as shown schematically in Figure 5.

Alternatively, a double ring peptide comprising two complementary cyclic peptides where one targets one loop of a protein and the other targets a second loop within the protein can be synthesized by first adding Fmoc-glutamic acid(trt)OH to Novabiochem TGR resin followed by Fmoc-Asp(Dde) and then synthesizing the complementary peptide using Fmoc-protected amino acids with tBu and Boc side chain protection as before. Prior to Fmoc deprotection of the final amino acid, Dde is removed as previously. Then Nα-Fmoc is removed and cylization is accomplished by coupling (as in SPPS) the Nα amine to the unprotected Asp COOH. The second cyclic peptide is synthesized on Novabiochem amino TGR resin using Asp(Dde)OH as the first amino. This is removed from the resin by 25% TFA, filtered, lyophilized and redesolved in DMF. The trt protection of Glu for the resin bound cyclic peptide is removed with 1% TFA as per Novabiochem protocol (see Catalog). The free cyclic peptide is then coupled (via its amide functionality) to the resin-bound, unprotected Glu (its COOH side chain) by DIC or other coupling, as previous, and the double ring peptide is released with 95% TFA, as are its side chain protecting groups. Purification and lyophilization, etc. as before.

The folding assay to be discussed later will be sufficient to test and evaluate the complementary peptide lead compounds, however they can also be evaluated based on their ability to bind to their target peptides. This optional test can be done while peptides are linked to the resin support (on bead binding assay); i.e., for branched or fusion peptides, or for resin-attached cyclic peptides; or free peptides will be coupled to an activated sepharose resin (e.g., Pharmacia) for the binding assay. The resin chosen must not bind any of the side chain amino acids of the peptide, because the binding assay requires that side chains be available for mediating complementary peptide association; e.g., peptide coupling via a carboxylic acid moiety can be done in the absence of aspartic and glutamic acid in the original peptide sequence. Water insoluble peptides can be dissolved in organic solvent and coupled to epoxy-activated resin (Pharmacia), and then assayed in aqueous buffer. In some cases peptides may be modified to ensure an appropriate binding mode to these resins.

The resin should be able to swell in water (for on-bead binding assays). This can be done by synthesizing the complementary peptides on Novabiochem amino TG resin and utilizing a binding assay in aqueous buffer (see **reference** 2 for details including blocking protocol) or as described above by attaching free peptide to a sepharose resin. Where complementary peptide is bound to a resin, the target protein's ability to bind the resin-bound peptide will be tested by using free target protein or part of the target protein (containing the sites targeted by the complementary peptide) which has been radiolabeled and unfolded, as previously described (SDS denaturation followed by dilution in Triton-X 100 containing buffer). Approximately 0.1 ml resin is exposed to 1μM or greater target protein concentration for the incubation times noted in reference. The resin is washed 3 to 5 times in same aqueous buffer containing 1% Triton-X 100 and radioactivity quantitated by scintillation counting. Specific activity of labeled protein based on protein content is used to measure mg bound protein to moles lead (peptide) compound (resin is saturated with peptide) and compared to control (non-target protein "probe" using same mg/unit gel as target).

Alternatively, (also part of optional binding test) and for cyclic peptides that require release from resin for cyclization, the target polypeptide (containing both target sequences but unlabled) is bound to Affi-gel or other sepharose resin, as above (see Bio-Rad instructions) (amount should be less than 30mg/ml protein, and placed in a column. (Control = non-target protein as above) The bound protein is then denatured by equilibrating the column with 2-3%SDS in aqueous buffer and allowing it to incubate at 30-37°C for 1 to 4 hours. The column is then washed by passing through 5 volumes of an aqueous buffer containing 0.1% Triton-X 100. Free complementary peptide (at 1/10 or less the concentration of bound protein) is then passed into the column (in same aqueous buffer as wash), flow is stopped (to retain peptide) and column is allowed to equilibrate for 2 hours (This may also be done by resuspending resin with added peptide and allowing resin to settle). Then the column is washed with 10 column volumes of same buffer (without peptide). Wash and initial flow through are collected, pooled and buffer is exchanged by gel filtration for either 100mM ammonium bicarbonate or 10% acetonitrile (depending on peptide solubility), and lyophilized. Residue is dissolved in 10% acetonitrile and run on reverse phase HPLC (C4 or C8). Peptide fractions are identified by optical

density (OD_{280nm}) and quantified by curve integration (exact procedure varies with equipment). Percent binding to receptor is calculated by comparing unretained peptide from gel-bound target protein to same from non-target protein: % binding = amount unretained peptide from protein control column, minus amount unretained peptide from target protein column, divided by amount unretained peptide from protein control column, multiplied by 100.

Optimization of Complementary Peptides

A complementary peptide that has been shown to inhibit the folding of a target protein may be optimized by synthesizing higher order peptide libraries based on complementary peptides sequences shown to inhibit folding of a target protein (see Multiple Peptide and Non-Peptide Libraries, chapter 4. G. Jung, editor. VCH 1996). Here screening for an optimized library of peptides is based on screening for enhanced affinity of peptides for a target sequence. This will in many instances positively correlate with increased folding inhibitory activity.

Peptoids

Peptoid combinatorial libraries have been prepared in which members contain side chain moieties common to the proteogenic amino acids. Hence these are like "peptides" in which the backbone chemistry has been altered. The binding interactions of complementary peptides are mediated by side chain interactions; therefore peptoids and related oligomers are likely to possess similar properties. Thus peptoids may be synthesized which contain the same or similar side chain units of a complementary peptide, e.g., as developed earlier as a folding inhibitor. Such a peptoid may also function as a folding inhibitor. Additionally, the synthesis and screening of combinatorial peptoid libraries may be employed as previously described for peptides. (See Multiple Peptide and Non-Peptide Libraries, pp. 387-404 Jung, editor. VCH 1996)

Beta Sheet Mimetics

Where a target protein domain contains a beta sheet structure formed by 3 parallel or anti-parallel beta strands, a synthetic peptide consisting of the sequences of the two strands that flank the central strand is synthesized with the following

modification. The two peptide sequences are either synthesized as a branched peptide, as previously defined, or as a cyclic peptide where the two peptide sequences are joined through glycine residues as was previously described. Alternatively a structure is synthesized that is similar to the "host" structure described in LaBrenz, S.,R., Kelly, J.,W. (1995) J. Am. Chem. Soc. 117, 1655-1656. In this case the peptide component of the structure consists of the above mentioned flanking beta strand peptides. An additional lead compound will consist of a single peptide comprising the sequence of the central peptide in the above described beta sheet.

Targeting the Molten Globule

The protein molten globule is a folding inhibitor target. Proteins go through a stage in folding termed the "slow phase." During this phase of folding, proteins are compact, contain native-like secondary structure, but do not have stable tertiary structure. Acquisition of stable native tertiary structure (completion of native folding) requires that the folded units of secondary structure associate in the proper orientations and make the contacts that result in the characteristic packing and folds of the native protein. Interfaces formed by peptides involved in secondary structure are targets for the class 2 inhibitors previously discussed.

Additionally, the fluorescent dye ANS binds to proteins in the molten globule state and has much less affinity for either completely unfolded or native, folded proteins. Binding is otherwise non-selective. ANS provides a scaffold which may be tailored and structurally appended, e.g., through combinatorial chemistry, to yield selective folding inhibitor leads that bind to the molten globule state of a target protein. Any other dye which similarly binds in the molten state may likewise be used.

Computer Aided Design -Class 2 Folding Inhibitors
(Peptides, Peptidomimetics, and Other Inhibitors of Protein-Protein Interactions and Interfaces)

Protein-protein interactions, such as the association of oligomeric subunits to form a large protein complex, as well as the association of two interacting proteins, can be inhibited in several ways. One means is to use a peptide (inhibitor) whose

sequence corresponds to a peptide segment within a larger protein, located at the surface of that protein, where the surface mediates contact and binding to another protein. If the noted region at the surface interface contributes critically to protein-protein binding, the free peptide fragment can prevent interaction of the larger protein units by competing for a binding site at the interface. Such peptides have been used to inhibit viral replication, immune recognition, and to prevent binding of fibronectin to a variety of receptor proteins. Cyclic (or circular) versions of such peptides often have enhanced potency because conformational constraint, produced by cyclization (circularization), reduces the entropy of binding and causes the cyclic peptide to spend more time in a biochemically relevant conformation.

Protein-protein interfaces are also generated within single polypeptide chains through the contacting of different folded regions of the polypeptide. This often occurs between domains in multidomain proteins and between subdomains (secondary structural motifs) within a domain. No agents that inhibit such contacts have previously been produced, however.

In eukaryotic cells domains fold sequentially and co-translationally. And even when such proteins fold spontaneously (and globally) *in vitro* (or post-translationally in bacteria) individual domains fold independently and their interfaces form later through tertiary structural rearrangement. A similar process holds for subdomains, exemplified by the fast (acquisition of secondary structure) and slow (acquisition of stable tertiary structure) phases of folding (i.e., upon spontaneous folding secondary structures form first and rapidly, followed by their slower rearrangement in space relative to one another to form stable tertiary structure).

Due to the independent folding of protein domains and to the co-translational folding of multidomain proteins, especially in eukaryotes, the opportunity exists for binding a ligand to a folded N-terminal domain before a C-terminal domain has been synthesized and folded or to one or another independently folding domains. If native folding requires that N- and C-terminal domains form an interface, then binding a ligand to an interfacial surface of one or both of the domains can inhibit subsequent folding through steric effects,through inhibition of peptide-peptide attraction (where the attraction occurs naturally between peptide sequences within the protein based on natural complementary hydropathy or other attractive

properties, or by inhibiting affinity-producing interactions that stabilize contacted domains relative to one another. (The previous discussion also pertains to subdomain-subdomain interfaces.) Where such interactions are known or predicted, selection of potential target sites can be narrowed.

Such a ligand may be a peptide whose sequence is found in a region of the polypeptide chain at the domain-domain interface or by a non-peptide ligand (small organic molecule) that binds specifically to a pocket or other structure formed by the surface of one or both interacting domains. (The inhibition of CD4 binding to MHC class II utilizes this type of ligand which was discovered through a computer screening approach.)

In general, the ligands discussed above (peptide or non-peptide) will be most effective if their target consists of the surface formed by the N-terminal domain, because in a co-translational folding system the N-terminal domain folds first and exists for a time without the contacting C-terminal domain. During this time which may be more than 1 minute, the domain is accessible for binding, thus allowing for inhibition of the interface contacts between the N- and C- terminal domains. In contrast, an agent which binds only to the C-terminal domain will have less time to bind before folding of the protein is completed but may still inhibit formation of the native protein interface or where folding completely post-translational.

Hence, this approach involves discovery of a ligand which binds to an interfacial surface joining two domains in a multidomain protein prior to complete folding of the protein. Such a ligand represents a Class 2 folding inhibitor and will function in the cell either during co-translational folding, post-translational folding, or refolding. This is best accomplished by a lead compound that binds the N-terminal domain on its domain-domain interface. There are two types of such leads: 1) leads that bind pockets and surface features, and 2) leads that are peptides or peptidomimetics. The first type of ligand will bind a pocket on the interfacial surface of this domain --a surface that in the native protein is generally sequestered from the solvent and which forms a part of the interface joining the two domains. Discovery of such a lead compound is accomplished by first using the known structure of the target protein, e.g., based on X-ray crystallography, and in some cases NMR solution structure, and defining the polypeptide sequences comprising each of the interfacial domains. A datafile consisting of the chosen N-terminal

protein domain is produced based on structural coordinates (saved from complete structural datafile; e.g., Brookhaven PDB). This becomes input for a program, e.g., DOCK, which derives a set of descriptors defining pockets and other structural features on the domain interfacial surface. At this point DOCK scans a chemical database (The Fine Chemicals Directory) whose structures have been encoded by the program CONCORD, to determine which compounds fit the surface feature of interest, and which will therefore have the potential to act as a ligand. The flow chart shown in Figure 6 summarizes the protocol. Specific references note computer programs and applications. Explanations of chart components is given below.

Target Protein

A target protein of known three dimensional structure is chosen where the protein contains at least two structural domains such that they are joined by a broad interfacial surface where a set of amino acids of one domain are in contact with a set of amino acids of the other domain. Glyceraldyhyde 3-phosphate dehydrogenase is an example. Domain-domain interfaces are noted in published reports of protein structures. Furthermore the amino acid sequences forming each domain must be known. For structurally defined protein this can often be ascertained by accessing a database of domain definitions, such as CATH (http://www.biochem.ucl.ac.uk/bsm/cath) or The Three Dee Database of Domain Definitions or a published structural report, as noted earlier.

Generating a Domain Datafile with Rasmol

The coordinates of a target protein are downloaded from a structural database (e.g., the Brookhaven PDB) and opened in the program Rasmol 2.5 or later (RasMol 2.5 or 2.6 Molecular Graphics Visualization tool. Roger Sayle, BioMolecular Structures Group, Glaxo Research & Development, Greenford, Middlesex, UK. June 1994). Amino acids corresponding to the sequence of an N-terminal domain forming the domain-domain interface are selected and saved as a PDB file (e.g., amino acids 1-150 for glyceraldehyde 3-phosphate dehydrogenase). As previously indicated, other molecular graphics programs can also be used. Those skilled in the art are familiar with how to select compatible programs.

Scanning Domain Surface Features with DOCK

The above file is used as input for the program DOCK 3.5. Although this program has been used for scanning discrete proteins or protein monomers, the domain datafile (constituting only a portion of a complete protein) will be used in the present invention and will be equivalent to other routine applications of this program in terms of method. Detailed methods concerning DOCK application are found in the following references: Shoichet, B.K., Kuntz, I.,D. (1991) J. Moll. Biol. 221, 327-346; GAO, S.,L.,J. et al. (1997) PNAS 94, 73-78; Ring, C.S. et al. (1993) PNAS 90, 3583-3587.

DOCK is used to define potential binding sites on the domain surface that forms the domain-domain interface. In the present invention, where two or more domains form such an interface, an N-terminal domain is chosen. A less than optimum (but potentially workable) condition results if a C-terminal domain is chosen (because ligands that bind a C-terminal domain will have less time available for binding during co-translational folding). Methods of characterizing and quantifying domain-domain interfaces are the same as those used for protein-protein interfaces and can be found in Jones, S., Thornton, J., M. (1996) PNAS 93, 13-20. Again, other compatible programs which can analyze protein surfaces and/or fit of molecular structures and/or surfaces can be used.

Scanning a Chemical Database for Lead Compounds

The program CONCORD (Concord: A Program for the Rapid Generation of High-Quality Approximate 3-Dimensional Molecular Structures. Austin: The University of Texas; St. Louis, MO: Tripos Associates) is used to transform the chemical formulas of compounds listed in a chemical database such as The Fine Chemicals Directory into structural formulations that can be used in DOCK to scan for ligands (lead compounds) capable of binding surface features (pockets), described above, forming part of the domain-domain interface. (See references cited above.) The potential ligands are then either acquired or synthesized and used as analytes to test folding inhibition. Also as indicated above, other compatible programs able to provide computer representations of molecular structure can be used.

SCANNING A VIRTUAL CHEMICAL DATABASE

The database of chemical compounds noted above does not have to represent a real collection of compounds. A very valuable alternative is to construct a virtual database, using computer methods, where the database is populated by representations of the compounds that could be synthesized during the creation of a combinatorial library (even though the library may not actually have been synthesized). As an example, consider the combinatorial library previously cited in Krchnak, V. et al. (1995) "Molecular Diversity, 1, 177-182. In this example a scheme is depicted (see fig. 3 in above reference) for each reaction used to generate the library and this includes a generalized formula for all compounds in the completed library. This type of information would of course be available for any type of combinatorial library. The general formula contains "R" groups denoting the chemical structures defining each unique building block. Hence a structural formula can be derived for each compound in the completed library by substituting all of the chemical moieties represented by "R" where all combinations of substitutions are listed. This set represents a virtual chemical database. It is treated as was the database noted previously (i.e., The Fine Chemicals Directory) by employing the programs CONCORD and DOCK as described (or other compounds providing similar analysis). Library members that are identified by DOCK as potential ligands can be quickly synthesized by (synthesis scheme) noting the specific building blocks and their locations within the library members. These library members (hits) can then be tested in actual assays and can also be used as scaffolds for the creation of biased libraries (which can also be tested in virtual form).

The use of a virtual chemical database can be used in all other applications of DOCK --e.g., where potential ligands are screened against a receptor-- and not just as a means to derive folding inhibitors or to inhibit protein-protein interactions.

Class 2 Inhibitors Consisting of Peptides and Peptidomimetics

The formation of a domain-domain interface can also be inhibited by introducing a free peptide containing the same amino acid sequence as a peptide forming part of the interface (part of the target protein). This is analogous to using peptides to inhibit protein-protein interactions through competitive binding (see

Tjernberg, L.,O. et al. (1996) J. Biol. Chem. 271, 15, 8545-8548, and Wild, C.,T. et al. (1994) *PNAS* 91, 9770-9774).

A peptide inhibitor can be designed by choosing a peptide from the target protein that is exposed to the domain-domain interface. Procedurally this does not differ from the derivation of peptide inhibitors discussed in the above references. However, instead of deriving the peptide from the solvent-exposed surface of a protein or protein subunit, it is derived from a site (the domain-domain interface) that is buried (not solvent exposed) in the native folded protein. The preferred formulation of said peptide in the present invention is a cyclic peptide. Here an additional 4 to 6 amino acids flanking the C- and N- terminal amino acids (e.g., 3 + 3 or 2 + 2) of the interface-exposed region are included during SPPS and cyclization is accomplished according to methods previously cited. (The additional amino acids may be part of the amino acid sequence of the target protein.) If the interface-exposed peptide forms a loop-like structure, then no additional amino acids are necessary and cyclization of the loop sequence will yield a potential lead compound.

Aptamers

Aptamers are RNA molecules or modified nucleic acids which through randomization and selection are able to bind specific targets, such as peptides, with very high affinity and specificity. Aptamers are generated against a peptide sequence within a target protein in the present invention. The aptamer is directed against a peptide located within the core of the target protein for the preferred embodiment (but other target sites are possible). The target peptide should preferably be between 6 and 16 amino acids long. Although a bivalent aptamer may be developed, the initial lead will consist of a single aptamer because it is possible to generate very high affinity aptamers against peptides. The protocol of choice is SELEX, described in Xu, W., Ellington, D. (1996) *PNAS* 93, 7475-7480. The resulting aptamer will be tested as an analyte in the folding assay to be described. Aptamers have also been covalently linked to cholesterol moieties to enhance intracellular transport, and this is also useful for developing aptamer folding inhibitors. Bivalent aptamers can also be made where a steroidal or other linker is used link two aptamers.

ASSAYS FOR FOLDING INHIBITION

After synthesis and preparation of the folding inhibitor lead compound, it will be used as an analyte to test its ability to recognize a target polypeptide and to inhibit the polypeptide's ability to fold in a physiologically relevant model system. (While analytes can be initially screened using a protein biosynthetic system in an initial stage, preferably potential inhibitors are initially identified using a binding assay, or computer-based binding identification method, or other convenient method.) This will preferably utilize an *in vitro* translation system. In general a eukaryotic target will be tested in a rabbit reticulocyte lysate translation system or in a wheat germ or oocyte system. A bacterial target protein can be tested in an *E. coli* lysate system (S30) for *in vitro* translation. All of these systems have been described and numerous examples (e.g., rabbit reticulocyte lysate, S30) are commercially available. (See, e.g., the following technical manuals from Promega Corp., which concern *in vitro* transcription and translation systems.

1. Rabbit Reticulocyte Lysate System Technical Manual, revised 8/96
2. Canine Pancreatic Microsomal Membranes Technical Manual, revised 7/92
3. E. coli S30 Extract System for Circular DNA, revised 11/97
4. Phage RNA Polymerases (SP6, T3 and T7), 3/97

All of the above either support the translation of RNA into protein or the transcription of DNA into RNA or both. The description below is exemplary and not limiting to the present invention.

The purified agent is first dissolved in a suitable solvent, such as DMSO, methanol, ethanol, or water (pH should be kept in the range 7-8.), as noted earlier, to produce a stock solution. Small amounts of such solvents do not, in general, perturb *in vitro* translation. The stock solution should be adjusted to a concentration at which dilutions of 20 or more fold produce concentrations of agent ranging from 1 to 100μM. A series of dilutions (final concentrations of analyte in the above range) will be made in a translation system such as nuclease treated rabbit reticulocyte lysate (Promega), to which an energy generating system, the complement of naturally occurring amino acids including one radioactive labeled amino acid, and other commonly used constituents such as salts, enzymes, etc. are added (as per above technical manuals).

In vitro translation of the target protein is then started either by introduction of RNA transcribed from a suitable vector (e.g., SP64) via commonly used protocols (see Promega manuals) containing cDNA for the target protein as an insert, or in a coupled transcription/translation reaction by introduction of DNA where the lysate is designed to support such a reaction (Promega). This will be done under routine conditions with respect to an optimized translation for the model protein (e.g., pH 7.4, at 30° to 37°C for 40 minutes).

The translations will be labeled with ^{35}S methionine or another appropriate labeled amino acid to allow for incorporation of the label into the translated polypeptide. Translations can be stopped, for example, by placing an aliquot, or the entire vessel containing the translation, on ice and by adding ice cold 2mM cycloheximide (for eukaryotic systems), 100μM chloramphenicol (bacterial translations), and 2mM methionine (unlabled for both) or other unlabled version of the amino acid used in labeling. As other alternatives, translation may be stopped with 0.01 to 0.1 mg/ml of Rnase A, or by dilution of aliquoted translation into SDS sample buffer, or into an ice cold, non-denaturing buffer (which can be followed by acid precipitation of protein).

The translation will then be subjected to sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions (e.g., with β-mercaptoethanol) known to anyone of ordinary skill in the art (see, e.g., protocols in above technical manuals). Samples are prepared in SDS sample buffer and heated according to common protocols (Promega). The resulting gels are stained, destained, dried and used to expose photographic film or a phosphoimager or other surface sensitive to ionizing radiation (techniques common to those of ordinary skill in the art). The signal thus generated is either observed visually and/or quantified through densitometry, fluorography, or a related method of measurement (see Netzer, W.J., Hartl, F.U. (1997) *Nature* 388, 343-349.)

Before conducting a folding assay, it should be ascertained whether the analyte and/or its solvent affect the translation quantitatively (i.e., the amount of translation product) and/or qualitatively (i.e., whether the translation product appears as a discernible band in its predicted position in a gel (denoting its molecular size) or whether multiple and/or off-sized bands occur). This can be done by using gel analysis as above to compare several (e.g., 5) *in vitro* translations in which

increasing concentrations of the agent are added (prior to initiation of translation) (e.g., 0M{solvent only}, 100pM, 1μM, 10μM, 100μM). Neither the solvent nor the agent should prevent translation of an analyzable product.

A folding assay, e.g., involving limited proteolysis, is then conducted to characterize the folded state of the translated protein in the absence of added analyte. This is called a titration and will be used as a standard to determine the effect that an analyte (lead compound) has on folding of the target protein. Proteinase K is a general protease (Boehringer-Mannheim) used for such assays. It does not exhibit significant amino acid sequence selectivity but rather requires only an exposed, flexible polypeptide backbone to catalyze proteolysis. As a result it does not digest folded protein (or large protein aggregates, even though these are not folded in the native sense) but does digest unfolded protein (that has not formed large aggregates) and exposed segments of the polypeptide chain within a folded protein. Thus, unfolded protein is digested to small peptides (that run off most gels and are thus not resolved) whereas folded protein is either left undigested or is digested to several resolvable and characteristic fragments which report on the native folded structure of the protein (See Fontana, A. et al. (1997) Folding and Design 2:R17- R26). Additionally, misfolded but aggregated protein is generally not digested. Alternatively, proteases such as chymotrypsin, trypsin and others may be used provided that one of these renders an analyzable signal (i.e., one that reports quantitatively on the foldedness state of the target protein).

The translation is stopped as above and a series of aliquots are diluted in buffer-containing vessels kept on ice (e.g., 3-5μl aliquots of translation diluted in 200-500μl of ice cold 20 mM Tris pH 7.4, 80 mM KoAc, 1-5 mM MgOAc). Alternative buffers may be chosen depending on compatibility with the protease and target protein. Then increasing concentrations of protease are added to each vessel. For proteinase K, vessel 1(V1) contains no protease, V2 contains 0.1μg/ml, V3 contains 1μg/ml, V4 contains 5μg/ml, V5 contains 10 or 20μg/ml. Then incubation on ice for 10 minutes after which PMSF is added to each vessel (to inhibit the protease) at a final concentration of 1mM (stock = 100mM in ethanol) and allowed to sit on ice for at least 5 minutes. Other proteases may use different inhibitors and incubation conditions. If the protein has been translocated to microsomes that have been added to the translation, then addition of protease is followed by 0.1%

digitonin unless a target domain is known to localize to the outer surface of the microsomal membrane (depending on the target protein, protease assays involving specific membrane associated proteins are also recorded in the prior art).

After the above limited proteolysis trichloroacetic acid (TFA) is added to each vessel to a final concentration of 10-15% and the vessels remain on ice for another 10 minutes. Precipitated protein is then sedimented by centrifugation (see Netzer, W.J., Hartl, F.U. (1997) Nature 388, 343-349.). Samples are washed with ice cold acetone and dissolved in SDS sample buffer and heated and resolved on SDS-PAGE as described above. A concentration of protease is chosen as a standard amount for later assays. This will be a concentration for which little or no change, compared to a lower concentration, was observed in the titration.

Folding Assay

Next a set of translations are conducted under identical conditions except for the concentration of added analyte --at least two translations contain no analyte (volume adjusted with translation buffer or analyte solvent, respectively). A second set of translations are conducted in parallel in which agent is added only after completion of translation (after stopping translation as above, and leaving on ice for at least 5 minutes; however incubation with subsequently added analyte is conducted at the translation temperature (e.g., 30-37°C); incubation time will equal the translation time used when analyte was added at start of translation (note, no further translation will occur here because translations were stopped with cycloheximide, etc. See above.). Samples are then returned to ice. Two equal aliquots (5µl each) of each translation sample are removed and added to separate vessels forming aliquot set A and aliquot set B, for each translation. Limited proteolysis as described above is conducted on set B samples using the protease concentration as determined previously and the samples are processed (TCA precipitated) and analyzed on gels as noted above. Set A samples contain no protease and are otherwise processed and analyzed in the same way. (Note: if the analyte at various concentrations is shown not to affect the yield or quality of the translation (prior to the folding assay), then a single sample may be used to represent set A.)

In general, a target protein that has been inhibited from folding will be completely or partially digested yielding no resolvable gel bands or will yield

numerous bands or a smear close to the bottom of the gel, and this will differ from the previously ascertained digestion pattern for native folded protein. (Exception: Aggregated protein may resist digestion. See below.) If less than 100% of the target protein has been inhibited from folding, then the full length product or the native pattern of limited proteolysis (in the absence of analyte) will be reduced in quantity proportionally and this can be measured by densitometry or fluorography or a similar method as noted. Thus a series of concentrations of analyte can be used and the potency of the analyte in inhibiting folding is measured as a function of its concentration and the degree to which the native pattern of limited proteolysis is altered (e.g., at 10μM analyte, native folded protein is 50% inhibited.)

Inhibition of folding can also result in protein aggregates that are resistant to protease. Under these circumstances the normal or native pattern of protease may be inhibited and the resulting gel analysis may show no difference between samples to which protease was absent or present. Aggregation, however, results from a failure of folding and therefore also constitutes a folding assay. In this case, in place of limited proteolysis, sedimentation (~50,000 x g, for 10 minutes at 4°C) of the samples of translation to which analyte has been added and to which analyte has not been added (as done above --but with protease left out) is conducted. Folded protein is soluble and does not sediment under these conditions. Large aggregates that are formed by unfolded or misfolded protein, sediment. Therefore a sedimentable translation product (target protein) indicates aggregation and misfolding; it is measured by comparing the quantity of sedimented target protein collected by washing the sedimentation vessel (after supernatant has been removed) with the quantity of protein present in the aliquot of translation before sedimentation. For example, if the sample volume is 20μl (before sedimentation), the sediment is dissolved and heated in 20μl of SDS sample buffer (or a calculation is made to relate the final volume of resuspended sediment and the volume of the original sample). Analysis is conducted on SDS-PAGE gel as before. (Variables below are measured in terms of quantified band intensity for given amount of sample.)

% Folding inhibition = 100 x (n μl sample for sediment (analyte treated sample) - n μl sample for sediment (analyte untreated sample))/n μl unsedimented sample.

Additionally, conformationally specific antibodies may be used to immunoprecipitate the variously treated samples, where immunoprecipitation either indicates a native fold or the absence of native fold.

In some cases biological activity can be assayed in the translation system and used as a measure of inhibition of folding because unfolded proteins will not be active in biochemical assays that measure enzyme activity, etc. Such an assay might be conducted in the case of HIV reverse transcriptase, for example. An RNA reporter template would be added to the sample at the completion of translation. Then a polymerase chain reaction would be conducted to amplify any reverse transcribed DNA. (PCR is required because the quantity of protein produced during *in vitro* translation is very small.)

In vitro translation systems (especially eukaryotic) such as rabbit reticulocyte lysate allow additional folding assays based on the ability to suspend protein elongation (of a nascent polypeptide) and essentially maintain the polypeptide bound to a ribosome. This is generally accomplished by removing the stop translation codon either in the mRNA or the DNA encoding the target protein. Usually this is accomplished by linearizing the plasmid DNA (a common technique for *in vitro* translation) through restriction enzyme digestion so that the cleavage site is located 5' (5 prime) of the stop translation codon and located within the coding sequence of the protein insert (or 3' [3prime] of the insert's start translation signal). Also the restriction site must not cut the promoter or separate the coding sequence from the promoter. Commencement of translation results in a truncated polypeptide that remains ribosome bound. (Optimization of translation temperature and salt concentrations can be employed to favor that a large portion of translating polypeptide chains remain bound to the ribosome.) If the truncation is in such a position that only a portion of a protein structural domain has been extruded from the ribosomal tunnel, then the resulting polypeptide will remain unfolded, as well as ribosome bound. This can be determined (assayed), for example, by limited proteolysis, as described. If the synthesized polypeptide consists of all or nearly all of a structural domain, then it may be made to fold into its native structure by inducing it to be released from the ribosome. Ribosomal release can be induced by incubation with various means, for examle, puromycin, chelation of magnesium (e.g., 10mM EDTA) or digestion with lmg/ml Rnase A. Additionally, the ribosomes

and their associated nascent polypeptides can be sedimented through a sucrose cushion, resuspended and polypeptide released as above. The folding assay would involve contacting the ribosome bound polypeptides with a test compound prior to release of the polypeptide from the ribosomes followed by release of the polypeptide and followed by assay of the amount of folded protein as described. Such a folding assay would require that said released polypeptide was foldable or that a target domain was complete enough to be foldable. In the absence of foldability (e.g., due to the shortness of a truncated polypeptide) the test compound could be assayed for binding to the truncated polypeptide but this is less than optimal in the present invention

Folding Assays Utilizing Random Screening

Folding inhibitors can also be discovered by random screening of chemical libraries. This could take the form of adding an unknown analyte to a biosynthetic or chaperone (see below) based assay, and then testing for folding inhibition. This may be less efficient in the context of the present invention because of the difficulty of adapting high throughput screening to such assays. The screening and deconvolution of pools of test analytes (mixtures of compounds from, e.g., a large chemical library) (see Houghten, R.A. et al. (1991) Nature 354:84-86) represents another method that can be applied to the discovery of folding inhibitors. However, this approach is not presently preferred as the method is cumbersome and there are significant variables in a biosynthetic assay or in a multimolecular (chaperone based) assay. For example, in a biosynthetic assay mixtures of analytes could perturb protein synthesis. This can be controlled however by determining that the fidelity of translation (e.g., in an *in vitro* translation system) has remained intact, as described earlier. However, there is a high likelihood that some compounds in a large mixture will prevent biosynthesis. In a chaperone-based assay, perturbation of chaperone ("foldase") activity by compounds present in a mixture of test analytes may also occur, but could be controlled by employing alternative substrate proteins as controls.

Preferably, random screening or other high volume screening is initially performed using a binding assay with a peptide probe or probes corresponding to a target protein. A folding inhibition assay can then be used to test or confirm the inhibitory activity of compounds which bind to probe.

Refolding of Already Synthesized Proteins

Other kinds of folding assays have been described and can be utilized in appropriate embodiments. Examples consist of rapidly diluting unfolded protein into a native buffer and assaying subsequent folding 1) biochemcially (enzymatic activity), 2) through light scattering, 3) specrophotometrically, 4) fluorgraphically, or 5) through the ability of a folded protein to bind ligand.

Scriptgen, Inc. has developed the Atlas™ system which utilizes a previously synthesized and purified protein in a reversible unfolding assay (US Pat. No. 5,679,582, which is hereby incorporated by reference in its entirety). The protein is exposed to denaturing conditions and then allowed to refold in the presence of a test analyte. If the analyte binds the folded protein and is present before denaturation, it can stabilize the folded form of the protein against unfolding (this reports a ligand of the folded protein). If the analyte binds an unfolded form of the protein it may prevent or inhibit refolding. However, the application of this assay to discover folding inhibitors requires that the protein be capable of reversible unfolding. Also, detection of a positive (an analyte that inhibits refolding) generally requires that the analyte must bind the protein after denaturing conditions are removed (because these would inhibit binding) and that it therefore bind a collapsed form of the "refolding" protein because extended (fully unfolded) forms of the protein would only be available for few milliseconds or less.

Protein folding consists of a rapid folding phase followed by a slower folding phase. For example, if a complete polypeptide chain is held in an unfolded state either by denaturant or by its attachment to a ribosome (in the context of its synthesis) or to chaperone, and then released into a folding environment (non-denaturing solvent), the rapid phase of folding (occurring within milliseconds or less) involves the collapse of the fully unfolded polypeptide into a compact intermediate which does not maintain stable tertiary structure but does have most of its hydrophobic residues buried within its core region. The subsequent slow phase of folding consists of the formation of stable, native tertiary structure which can take from several seconds to minutes or longer. Thus, in an *in vitro* refolding assay, a polypeptide placed in non-denaturing (refolding) solution exists in a completely unfolded state for only a very short time. In contrast, this completely unfolded form of a protein exists for a much longer time in the living cell in the absence of

denaturant. That is because a polypeptide chain remains in a completely unfolded (or non-collapsed) form (e.g., near random coil) during its association with ribosomes (which may last for minutes in humans) or chaperones (which may last even longer).

The methods of the present invention provide for the discovery of compounds that inhibit protein folding by binding to the completely unfolded form (i.e., random coil, non-collapsed forms) of a target protein which is made possible by the use of a biosynthetic assay and/or chaperone. Many compounds capable of binding this unfolded form will not be capable of binding a compact intermediate in which solvent has already been excluded from a core region. Furthermore, compounds capable of binding a fully unfolded form may not do so in an *in vitro* refolding experiment because the unfolded form is not present long enough in the absence of denaturant (note: denaturant will generally abolish many types of ligand binding).

Additionally, since folding inhibition antagonizes the target protein, biological or *in vivo* assays (e.g., DNA synthesis, tumor growth) could be used to measure a biological response. Therefore, any biological, biochemical or functional assay which requires synthesis of a target protein is a potential folding assay.

Additionally, recovery from heat shock or stress (measured by recovery of a biochemical activity otherwise sensitive to heat shock associated with the target protein, could be used as an assay because such recovery requires protein refolding). Cells are pulse labeled and chased (where folding will be assayed by proteolysis or by cellular degradation) and then subjected to an interval of heat shock (in the presence or absence of a folding inhibitor test analyte), followed by sampling at several time points after removal of heat shock, with SDS PAGE analysis. This may alternatively be done without labeling (pulse) where a biological activity of a target protein is assayed. Folding inhibitors that are also inhibitors of refolding, can thus be used with heat shock in multimodal therapies.

Folding Assay based on Chaperones

Chaperones bind unfolded forms of proteins. Chaperones of the Hsp70, Hsp90 and Hsp40 classes bind relatively unfolded forms of proteins resembling random coils and also bind short peptides. Chaperonins (chaperones of the Hsp60

class) bind unfolded proteins which may be in a slightly more compacted state than when bound by chaperones of other classes. Within cells, unfolded proteins may be bound by several chaperones at the same time. For example, in *E. coli* unfolded proteins may be bound by dnaK (Hsc70) and dnaJ (Hsc40). In eukaryotes, nascent polypeptides may be bound by Hsc40 and Hsc70. Various signal transduction proteins are bound by Hsc90 and a variety of auxiliary chaperones. Proteins translocated to the endoplasmic reticulum may be bound, and stabilized in an unfolded state by numerous ER resident chaperones. Comparable situations exist for other organelles.

In addition to existing inside cells, chaperone-protein complexes (unfolded substrate protein bound to chaperone) can be easily produced *in vitro*. In general this is accomplished by denaturing a purified protein and rapidly diluting it into a buffer containing chaperones of interest (generally in the absence of ATP and/or Mg++). In these systems the diluted protein binds to chaperones and forms stable complexes. (Alternatively target protein-chaperone complexes may be isolated from cells or cell lysates.) In the presence of ATP and Mg++ chaperone/protein complexes also form but dissociate quickly. While bound to chaperone, a substrate protein is in an unfolded state. For most such complexes the bound protein is released by adding ATP and Mg++ to the solution. Most chaperones are ATPases and readily hydrolyze the added ATP. By binding to and/or hydrolyzing ATP the bound substrate protein is released into the bulk solution in the case of non-chaperonin chaperones (e.g., Hsp70), where it either folds, aggregates, or rebinds chaperone in an unfolded state. In the case of chaperonins (specifically the chaperonin GroEL) protein that fails to fold after release from the GroEL cavity does not enter the bulk solution but tends to rebind the cavity in an unfolded state.

The ability of a protein to fold after release from chaperone can generally be measured quite easily. This involves measuring folded protein after addition of ATP/Mg++ as has been described in the art (see references below) and utilizing any of the variety of folding assays (to discern folded from unfolded protein) as noted elsewhere in this disclosure.

In the case of protein bound to Hsp90 or to an Hsp90 containing complex (as has been described elsewhere), in many instances the bound portion of the protein is an unfolded domain whereas other domains of the protein may be folded. This is

generally the case for various signal transduction proteins whose regulation and/or maturation are mediated by Hsp90. Other proteins that bind Hsp90 during stress may be completely unfolded. In either case, release of protein from Hsp90 does not readily occur by addition of ATP/Mg++. However, if an analyte being tested as a folding inhibitor binds a protein bound to Hsp90 it can be inferred that such an analyte may be capable of binding the protein while it is in an unfolded and physiologically relevant state. Hence, such an analyte may be a folding inhibitor. So, binding to Hsp90 (without subsequent release) may also constitute a folding inhibition assay.

Binding here can be assayed by co-immunoprecipitation of analyte and Hsp90 complex (e.g. using anti-Hsp90 antibody), co-sedimentation of analyte and Hsp90 in a sucrose or other gradient or chromatographic medium, or co-migration of analyte and Hsp90 in gel filtration or gel electrophoresis, or by covalent cross linking of analyte and Hsp90 or Hsp90 complex and subsequent isolation or chromatography. Binding to chaperone can be ruled out by using a control consisting of a different substrate protein. So specific binding to an unfolded target protein can be measured. Such binding assays can be utilized for other chaperones and chaperone systems without release of bound target protein, however, release and subsequent folding assay is preferred when such release is feasible.

Folding assays designed to discover folding inhibitors will in general take the following forms. A target protein will be unfolded by dissolving in denaturant such as 8M urea or 6M guanidinium hydrochloride. In some instances a target protein may be denatured by heat and this may be carried out in the presence of chaperone. Otherwise the dissolved, denatured target protein will be rapidly diluted by a factor of 100 into a non-denaturing buffer containing approximately two molar equivalents of at least one chaperone, such as GroEL or Hsc70. This buffer will either not contain Mg++ and/or ATP. A Mg++ chelator such as CDTA or EDTA may also be present; if a chelator is present then the initial solution may contain Mg++ but the amount of chelator should be adjusted so all or most of the Mg++ is predicted to be chelated.

Several aliquots (samples) of the mixture will be made and a test analyte representing a potential folding inhibitor will then be added to all but one of the samples (controls will in general also be added to other samples) at various

concentrations below and above the concentration of previously added target protein. (Highest analyte concentrations may be 10 or more fold greater than the concentration of the target protein.) This mixture is allowed to stand at room or other temperature (e.g., 30°-37°C) for around 5 minutes to 1 hour. Then Mg++ and/or ATP are added so that the resulting concentration of each are around 5mM (if chelator is present then an amount of Mg++ should be added so that the final predicted concentration of Mg++ is 5mM). (Lower or higher concentrations of Mg++/ATP are also permissible.) Negative controls will consist of chaperone substrate proteins other than the target protein.

The addition of Mg++/ATP will cause release of the target protein which will then have an opportunity to fold (or aggregate or rebind to chaperone in an unfolded state). Folding can then be assayed by any of the means previously noted and comparison of each sample (containing different concentrations of test analyte or no test analyte) can be made to assay folding inhibition as a function of analyte concentration. In addition to previously mentioned folding assays, unfolded protein may rebind to chaperone. The amount of such rebound target protein can be assayed and also used as a measure of unfolded protein (folding inhibition) by adding a Mg++ chelator to the samples after incubation with test analyte. Any chaperone bound target protein will then be stabilized on the chaperone in an unfolded state and the amount of such complex can be measured by, for example, gel filtration of the chaperone/target protein complex. The greater the amount of chaperone bound target protein (or a lesser amount of bound protein in the presence of aggregated target protein) the greater the degree of folding inhibition.

The final concentration of target protein used in a chaperone system will, in some instances, be less than 1μM so that aggregation of the protein is avoided. In various instances this will allow a more robust assay of folding inhibition because the confounding factor of aggregation will not be operating. In other instances a higher concentration of target protein will be used. In these instances folding can still be assayed, as well as aggregation which can also be used to assay folding inhibition.

Folding assays can also be carried out using chaperone trap proteins (e.g., GroEL-trap). These proteins are mutated forms of specific chaperones, including the chaperonins. These "traps" or "trap proteins" share the property of binding

essentially irreversibly to unfolded substrate proteins. In other words, unfolded proteins may bind the trap but are not subsequently released as with normal chaperones. In an *in vitro* assay, a trap can be added after (but usually not before): 1) the contacting of an unfolded protein with a test compound, and 2) the subsequent attempted refolding of the target protein is produced. Protein inhibited from folding by the test compound will bind trap and can be assayed by detecting the protein associated with trap by gel filtration, co-immunoprecipitation, or similar means. Folded protein will not bind trap. The quantity of the bound protein is a measure of the quantity of folding inhibited protein, preferably compared with suitable controls.

A preferred embodiment using chaperone traps, which include trap versions of the chaperonins, is carried out where the trap is contacted to a protein biosynthetic system (especially a eukaryotic *in vitro* translation system such as rabbit reticulocyte lysate) under protein synthesis conditions. The target protein will be synthesized and will not bind to trap (e.g., where GroEL-trap is added to a eukaryotic protein biosynthetic system) if the protein undergoes normal *de novo* folding. If the protein is inhibited from folding by a test compound added to the biosynthetic system, the folding inhibited protein will bind to trap and can be detected and measured as previously described. Hence, binding to trap during *in vitro* translation is an assay for folding inhibition and should be carried out with suitable controls. Additionally, similar folding assays can be carried out within living cells in which a trap chaperone or other trap protein has been introduced, e.g., by gene transfection or by physical delivery, such as electroporation or microinjection. The trap chaperone will generally be present before test compound is contacted to the biosynthetic system and before synthesis and attempted folding of the target protein (in contrast to *in vitro* refolding assays). It is also expected that non-chaperone proteins will be engineered and adapted to function as "traps" and may be used in place of chaperone traps in folding assays. Thus, the term "trap" or "trap protein" refers to a protein which preferentially binds unfolded proteins and will remain bound under conditions where normal chaperones or chaperonins would be released, preferably under conditions which would disrupt the normal folded structure of the protein. Preferably the binding is irreversible Preferably, such trap protein are mutated chaperones or chaperonins, e.g., as described in the art.

The following references detail some assays that can be used in this embodiment. In each case, however, the addition of a test analyte (for folding inhibition) will be expected.

Langer, T. et al. Nature (1992) 356(6371) 683-689; Martin, J. et al. Science (1992) 258(5084) 995-998 and Nature (1993) 366(6452) 228-233; Nimmesgern, E, Hartl, F.U., (1993) Febs Letters 331(1-2) 25-30; Schneider, C. et al. (1996) PNAS 93 (25) 14536-41.

Protein Unfolding

Additionally, some folding inhibitors will denature folded proteins. This results because folded proteins are in rapid equilibrium between folded and unfolded forms. The unfolding can be large scale, involving essentially the entire protein and/or can involve only local structure with the protein. A compound able to bind an unfolded form of a protein may shift this equilibrium towards unfolding. This has been demonstrated, for example, with an antimyoglobin antibody. It has also been demonstrated for other antibodies (Wien, M.W., et al. (1995) Nature Structural Biology 2, 232-243.) Hence, a folding or functional assay can be produced by incubating an analyte with a folded protein and then employing a folding, biochemical, or biological assay as previously noted. Unfolding can also result from binding of a ligand to the surface of a folded protein, where the ligand alters the environment of the protein, causing the folded state to be disrupted. This characteristic can also be utilized in an assay for protein folding inhibitors. The disruption can be due to any of a number of different parameters affecting folding equilibrium and/or structure. For example, a ligand can introduce strain in the folded structure, destabilizing the native folded form, or a ligand could change the electrostatic and/or hydrophobicity characteristics of the local environment.

Preparation & Administration of Protein Folding Inhibitors

For the treatment of patients suffering from a disease or other condition in which the inhibition of a protein is desired, the preferred method of preparation or administration will generally vary depending on the type of compound to be used. Thus, those skilled in the art will understand that administration methods as known in the art will also be appropriate for the compounds of this invention.

The particular compound that exhibits protein folding inhibitor activity can be administered to a patient either by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s). In treating a patient exhibiting a disorder or condition of interest, a therapeutically effective amount of an agent or agents is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of one or more symptoms or a prolongation of survival in a patient.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures and/or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds which exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by HPLC or other means appropriate for detection of the particular compound.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition (see *e.g.* Fingl et. al., in The Pharmacological Basis of Therapeutics, 1975, Ch. 1 p.1).

It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions, or to systemic thiamin deficiency. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administered dose in the management of the disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for

example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and response of the individual patient. A program comparable to that discussed above also may be used in veterinary medicine.

Depending on the specific conditions being treated and the targeting method selected, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in Alfonso and Gennaro (1995). Suitable routes may include , for example, oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, or intramedullary injections, as well as intrathecal, intravenous, or intraperitoneal injections.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular those formulated as solutions, may be administered parenterally, such as by intravenous injection. Appropriate compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently

delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions, including those formulated for delayed release or only to be released when the pharmaceutical reaches the small or large intestine.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such

as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

Delivery of Nucleic Acid Molecules

Certain of the inhibitor compounds of the present invention may be nucleic acid molcules or may be encoded by nucleic acid molecules and expressed within cells following administration. Introduction of nucleic acid sequences encoding and expressing an inhibitor compound or a plurality of such agents into target cells by gene delivery or gene therapy provides a means to express protein folding inhibitors in the targeted cells. Some approaches and results in gene therapy were reviewed in Miller (1992): Human gene therapy comes of age. *Nature* 357:455-460); Eck and Wilson (1996). Gene-based therapy. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics, Ninth Edition* (ed. Hardman, J.G. and Limbird, L.E.), pp. 77-101. McGraw-Hill: New York); and Anderson (1998). *Nature* 392 (suppl.). 25-30). The methods described in those references can be utilized in the present invention.

Many cell types have been targeted for gene therapy, including lung epithelium, transplanted bone marrow cells, skin fibroblasts, and so forth (Watson et al. (1992). *Recombinant DNA, 2nd edition*. Scientific American Books: New York, Chapter 28; Eck and Wilson, 1996).

Targeting has been accomplished by diverse means, including direct injection, aerosols into the lung, use of a virus with a targeting ligand in its envelope (e.g., Han et al., 1995), addition of a specific ligand to the DNA (Lu et al., 1994), receptor-mediated uptake (Perales et al., 1994), liposome encapsulation (Vieweg et al., 1995), or even systemic administration of a gene that is expressed only in the target tissue (Arteage and Holt, 1996; Lee et al., 1996). The methods listed will allow targeting of suitably engineered genes encoding protein folding inhibitors.

Along with the various methods of targeting, a number of different delivery methods can be used. A variety of such delivery methods are known in the art; some methods of delivery that may be used include:

a. complexation with lipids,
b. transduction by retroviral vectors,
c. localization to nuclear compartment utilizing nuclear targeting sites found on most nuclear proteins,
d. transfection of cells *ex vivo* with subsequent reimplantation or administration of the transfected cells,
e. a DNA transporter system.

A nucleic acid sequence encoding a protein folding inhibitor may be administered utilizing an *ex vivo* approach whereby cells are removed from an animal, transduced with the nucleic acid sequence and reimplanted into the animal. For example, the liver can be accessed by an *ex vivo* approach by removing hepatocytes from an animal, transducing the hepatocytes *in vitro* with the nucleic acid sequence and reimplanting them into the animal (*e.g.*, as described for rabbits by Chowdhury et al, *Science* 254: 1802-1805, 1991, or for humans by Wilson, *Hum. Gene Ther.* 3: 179-222, 1992) incorporated herein by reference).

Exogenous cells can also be used. In this approach a vector encoding a protein folding inhibitor, is inserted into cells. If desired, the cells can then be grown in culture. The cells carrying the vector are then delivered into the animal to

be treated. Preferably the cells are targeted or localized to the locality of the targeted cells.

Many nonviral techniques for the delivery of a nucleic acid sequence encoding a protein folding inhibitor into a cell can be used, including direct naked DNA uptake (*e.g.*, Wolff et al., *Science* 247: 1465-1468, 1990), receptor-mediated DNA uptake, *e.g.*, using DNA coupled to asialoorosomucoid which is taken up by the asialoglycoprotein receptor in the liver (Wu and Wu, *J. Biol. Chem.* 262: 4429-4432, 1987; Wu et al., *J. Biol. Chem.* 266: 14338-14342, 1991), and liposome-mediated delivery (*e.g.*, Kaneda et al., *Expt. Cell Res.* 173: 56-69, 1987; Kaneda et al., *Science* 243: 375-378, 1989; Zhu et al., *Science* 261: 209-211, 1993). Many of these physical methods can be combined with one another and with viral techniques; enhancement of receptor-mediated DNA uptake can be effected, for example, by combining its use with adenovirus (Curiel et al., *Proc. Natl. Acad. Sci. USA* 88: 8850-8854, 1991; Cristiano et al., *Proc. Natl. Acad. Sci. USA* 90: 2122-2126, 1993).

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences into the targeted cell population (*e.g.*, tumor cells). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing sequences encoding protein folding inhibitors. *See*, for example, the techniques described in Sambrook et al. (1989) and in Ausubel et. al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, recombinant nucleic acid molecules encoding protein folding inhibitor protein sequences can be used as naked DNA or in a reconstituted system *e.g.*, liposomes or other lipid systems for delivery to target cells (*see e.g.*, Felgner et. al., *Nature* 337:387-8, 1989). Several other methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins.

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA (*e.g.*, a plasmid vector encoding a peptide or polypeptide into the nucleus of a cell, through a process of microinjection (Capecchi MR, *Cell* 22:479-88 (1980)). The DNA can be part of a formulation which protects the DNA from degradation or prolongs the bioavailability or the DNA, for example by

complexing the DNA with a compound such as polyvinylpyrrolidone. Once recombinant genes are introduced into a cell, they can be recognized by the cells' normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been used for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with $CaPO_4$ and taken into cells by pinocytosis (Chen C. and Okayama H, *Mol. Cell Biol.* 7:2745-52 (1987)); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu G. et al., *Nucleic Acids Res.*, 15:1311-26 (1987)); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner PL., et al., *Proc. Natl. Acad. Sci. USA.* 84:7413-7 (1987)); and particle bombardment using DNA bound to small projectiles (Yang NS. et al.. *Proc. Natl. Acad. Sci.* 87:9568-72 (1990)). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene (Curiel DT et al., *Am. J. Respir. Cell. Mol. Biol.*, 6:247-52 (1992)).

As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, or oligonucleotide or polynucleotide. Gene transfer can be performed in cultured cells or by direct administration into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cytoplasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene

transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer can be performed *ex vivo* on cells which are then transplanted into a patient, or can be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a protein folding inhibitor is provided in which the nucleic acid sequence is expressed only in specific tissue. Examples or methods of achieving tissue-specific gene expression are described in International Publication No. WO 93/09236, published May 13, 1993, filed November 3, 1992.

Oligopeptide and Polypeptide Chemical Derivatives of Protein Folding Inhibitors

The oligopeptides of this invention can be synthesized chemically or through an appropriate gene expression system. Synthetic peptides can include both naturally occurring amino acids and laboratory synthesized, modified amino acids.

Also provided herein are functional derivatives of a polypeptide or protein. By "functional derivative" is meant a "chemical derivative" of the polypeptide or protein. A functional derivative retains at least a portion of the function of the protein, expecially protein folding inhibitor activity.

A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein. Such moieties may improve the molecule's solubility, absorption, biological half life, and the like. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like. Moieties capable of mediating such effects are disclosed in Alfonso, R. and Gennaro, L.C. (1995): "Remington: The Science and Practice of Pharmacy, 19th ed.." Easton, PA: Mack Publishing Co.. Procedures for coupling such moieties to a molecule are well known in the art. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give

carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloro-mercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine- containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction carbodiimide (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are

deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking component peptides to each other or the complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimidate yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex. Moieties capable of mediating such effects are disclosed, for example, in Alfonso and Gennaro (1995).

Derivatives of Oligonucleotides

As described above, the inhibitors of the present invention may be aptamers, which generally are ribonucleotide sequences which bind to target peptide sequences, but which may also include modified nucleotides or nucleic acid analogs (for example to provide greater resistance to intracellular Rnases. Such molecules of the invention may be prepared by any method known in the art for the synthesis of RNA and related molecules. *See*, for example, Draper, PCT WO 93/23569. hereby

incorporated by reference herein. These include techniques for chemically synthesizing oligodeoxyribonucleotides well known in the art such as, for example, solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the aptamer or other nucleic acid molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, aptamer or other nucleic acid sequence cDNA constructs that synthesize aptamer or other sequence RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Various modifications to the DNA molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the use of phosphorothioate or methyl phosphonate rather than phosphodiesterase linkages within the backbone. Modifications may also be made on the nucleotidic sugar or purine or pyrimidine base, such as 2'-O-alkyl (*e.g.*, 2'-O-methyl), 2'-O-allyl, 2'-amino, or 2'-halo (*e.g.*, 2'-F). A variety of other substitutions are also known in the art and may be used in the present invention. More than one type of nucleotide modification may be used in a single modified oligonucleotide. In addition, portions of the aptamer or other nucleic acid sequence may contain one or more non-nucleotidic moieties.

Preferred oligonucleotide inhibitors include oligonucleotide analogues which are resistant to degradation or hydrolysis by nucleases. These analogues include neutral, or nonionic, methylphosphonate analogues, which retain the ability to interact strongly with complementary nucleic acids. Miller and Ts'O, *Anti-Cancer Drug Des.* 2:11-128 (1987). Further oligonucleotide analogues include those containing a sulfur atom in place of the 3'-oxygen in the phosphate backbone, and oligonucleotides having one or more nucleotides which have modified bases and/or modified sugars. Particularly useful modifications include phosphorothioate linkages and 2'-modification (*e.g.*, 2'-O-methyl, 2'-F, 2'-amino).

All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the invention pertains. All

references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The specific methods and compositions described herein as presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. For example, those skilled in the art will recognize that the invention may suitably be practiced using a variety of types of data within the general descriptions provided.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is not intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the

genus, regardless of whether or not the excised material is specifically recited herein.

In addition, where features or aspects of the invention are described in terms of Markush groups or other grouping of alternatives, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group or other group. For example, if there are alternatives A, B, and C, all of the following possibilities are included: A separately, B separately, C separately, A and B, A and C, B and C, and A and B and C. Thus, for example, for sets or types or ranges specified herein, the embodiments expressly include any subset or subgroup or individual item of those sets or types or ranges. While each such subset or subgroup or item could be listed separately, for the sake of brevity, such a listing is replaced by the present description. Thus, for example, for a range such as 4-16 amino acids, the present description expressly includes each individual point in that range (including the endpoints) as well as any subset of the range, and thus includes, for example, 4,5,6,7,8,9,10,11,12,13,14,15, and 16 amino acids, as well as subsets such as 4-8, 9-16, and 6-16 amino acids.

While certain embodiments and examples have been used to describe the present invention, many variations are possible and are within the spirit and scope of the invention. Such variations will be apparent to those skilled in the art upon inspection of the specification, drawings and claims herein.

Other embodiments are within the following claims.

CLAIMS

What is claimed is:

1. A method for identifying a protein folding inhibitor, comprising the steps of:

a) contacting a protein biosynthetic system under protein synthesis conditions with at least one test compound; and

b) determining whether said test compound increases the ratio of unfolded protein to folded protein, wherein an increase in said ratio is indicative that said test compound is a said protein folding inhibitor.

2. The method of claim 1, wherein said determining comprises comparing said ratio of folded protein to unfolded protein in the presence of said test compound to in the absence of said test compound.

3. The method of claim 1, further comprising contacting said protein with at least one chaperone protein.

4. The method of claim 3, wherein said protein is contacted with said chaperone protein prior to exposure to the presence of said test compound.

5. The method of claim 1, wherein said protein biosynthetic system is an *in vitro* system selected from the group consisting of a eukaryotic protein biosynthetic system and a prokaryotic protein biosynthetic system.

6. The method of claim 1, wherein said unfolded protein is detected using a method selected from the group consisting of a proteolysis with electrophoresis, aggregate sedimentation, binding to conformationally specific antibodies; binding to at least one chaperone protein, and determination of biological activity.

7. The method of claim 1, wherein said test compound is selected from the group consisting of a multivalent binding compound, a complementary peptide, an aptamer, and a small molecule.

8. The method of claim 1, wherein said test compound comprises a domain:domain interface sequence.

9. A method for identifying a protein folding inhibitor, comprising the steps of:

a) binding an unfolded protein with at least one chaperone protein to form a test combination;

b) contacting said test combination under non-denaturing conditions with a test compound;

c) releasing said at least one chaperone protein; and

d) determining whether said test compound increases the ratio of unfolded protein to folded protein, wherein an increase in said ratio is indicative that said test compound is a protein folding inhibitor.

10. The method of claim 9, wherein said unfolded protein is a previously synthesized protein subjected to unfolding conditions.

11. The method of claim 9, wherein said at least one chaperone protein is present in a protein synthetic system during synthesis of said unfolded protein.

12. The method of claim 9, wherein said determining comprises comparing the ratio of folded protein to unfolded protein in the presence of said test compound to in the absence of said test compound.

13. The method of claim 9, further comprising exposing said test combination to chaperone-binding stabilization conditions.

14. The method of claim 9, wherein said test compound comprises a domain:domain interface sequence.

15. A method for identifying a protein folding inhibitor, comprising the steps of:

a) providing a peptide, wherein said peptide is a potential protein-stabilizing peptide and does not require unfolding;

b) contacting said peptide with a test compound under non-denaturing

conditions; and

c) determining whether said test compound binds to said peptide, wherein binding of said test compound to said peptide is indicative that said test compound is a protein folding inhibitor.

16. The method of claim 15, wherein said peptide comprises a domain:domain interface sequence.

17. The method of claim 16, wherein said peptide is in an intact protein or in a portion of a protein comprising at least two domains.

18. The method of claim 15, wherein said peptide is in a folded polypeptide or protein.

19. The method of claim 15, wherein said peptide is in a protein or polypeptide bound to at least one chaperone protein.

20. The method of claim 15, wherein said test compound or a plurality of said test compounds are attached to a solid phase support.

21. The method of claim 15, wherein said peptide is attached to a solid phase support.

22. The method of claim 15, wherein said binding is detected using a method selected from the group consisting of detection of a label attached to said test compound, detection of a label attached to a molecule comprising said peptide, binding of an antibody to said peptide or protein, an electrophoretic mobility shift assay, and gel filtration.

23. The method of claim 15, wherein said test compound is selected from the group consisting of multi-valent binding compounds, complementary peptides, aptamers, small molecules, and members of a combinatorial library.

24. The method of claim 15 wherein said test compound comprises a domain:domain interface sequence.

25. A method for identifying a protein folding inhibitor, comprising the steps of:

a) contacting a folded protein or polypeptide with a test compound under non-denaturing conditions; and

b) determining whether the amount of unfolded protein or polypeptide is increased in the presence of said test compound.

26. The method of claim 25, wherein said unfolded protein is detected using a method selected from the group consisting of proteolysis with electrophoresis, aggregate sedimentation, binding to conformationally specific antibodies; binding to at least one chaperone protein, and determination of biological activity.

27. The method of claim 25, wherein said test compound is selected to bind at a domain:domain or sub-domain:sub-domain interface.

28. The method of claim 25, wherein said test compound is a domain:domain interface sequence.

29. A method for identifying a protein folding inhibitor, comprising the steps of:

a) contacting a protein or polypeptide with a test compound under protein-denaturing conditions, wherein said protein or polypeptide comprises a potential protein stabilizing peptide; and

b) determining whether said test compound binds to said peptide, wherein binding of said test compound to said peptide is indicative that said test compound is a protein folding inhibitor.

30. The method of claim 29, wherein said peptide comprises a domain:domain interface sequence.

31. The method of claim 30, wherein said peptide is in an intact protein or in a portion of a protein comprising at least two domains.

32. The method claim 30, wherein said test compound or a plurality of said test compounds are bound to a solid phase support.

33. The method of claim 29, wherein said peptide is attached to a solid phase support.

34. The method of claim 29, wherein said binding is detected using a method selected from the group consisting of detection of a label attached to said test compound, detection of a label attached to a molecule comprising said peptide, and binding of an antibody to said peptide or protein.

35. The method of claim 29, wherein said test compound is selected from the group consisting of multi-valent binding compounds, complementary peptides, aptamers, and small molecules.

36. The method of claim 29, wherein said test compound comprises a domain:domain interface sequence.

37. A method for identifying a putative protein folding inhibitor, comprising the steps of:

a) obtaining 3-D structural coordinates of a peptide or a plurality of peptides which form a structural domain or sub-domain of a protein;

b) identifying a surface of said domain or sub-domain which forms an interface with one or more other structural domains or sub-domains; and

c) docking a plurality of molecular structures on said surface to determine the quality of fit;

wherein identification of a molecular structure with a good quality of fit to said surface is indicative that a compound with said molecular structure is a protein folding inhibitor.

38. The method of claim 37, wherein said domain is an N-terminal domain of said protein.

38 The method of claim 37, wherein said docking comprises determining the putative molecular interations of each said molecular structure with said surface using computer calculation of the expected interaction free energy of said molecular structure with said surface.

40. The method of claim 37, wherein said surface is identified using an implementation of a DOCK program or a modification or derivative thereof.

41. The method of claim 37, wherein said molecular structure or structures is described using an implementation of a CONCORD program or a modification or derivative thereof.

42. The method of claim 37, wherein said molecular structures are structures from a virtual compound library, a real compound library, or both.

43. The method of claim 37, further comprising selecting a site or sites on said surface for said docking.

44. The method of claim 37, further comprising identifying said surfaces for a plurality of domains or subdomains of said protein.

45. The method of claim 37, wherein said domain or sub-domain is identified using 3-D coordinates for said protein or a portion thereof including said domain or sub-domain.

46. The method of claim 37, further comprising:

testing the ability of said putative protein folding inhibitor to inhibit protein folding utilizing an assay comprising a protein biosynthetic system or a chaperone protein or both.

47. The method of claim 37, wherein said molecular structures represent test compounds selected from the group consisting of peptides, aptamers, and small

molecules.

48. A method for identifying a putative protein binding compound, comprising the steps of:

a) identifying a protein surface and optionally selecting a site or sites;

b) docking a plurality of molecular structures from a virtual compound library on said surface or on said site or sites to judge the quality of fit of each said molecular structure; and

c) choosing putative binding compounds by selecting molecular structures predicted by step b) to have good quality of fit.

49. The method of claim 48, wherein said library is a virtual combinatorial library.

50. The method of claim 48, wherein said docking comprises determining the putative molecular interaction of each said molecular structure with said surface using computer calculation of the expected interaction free energy of said molecular structure with said surface.

51. The method of claim 48, wherein said docking comprises the use of an implementation of a DOCK program or a modification or derivative thereof.

52. The method of claim 48, wherein said molecular structure or structures id described using an implementation of a CONCORD program or a modification or derivative thereof.

53. The method of claim 48, further comprising selecting a site or sites on said surface for said docking.

54. The method of claim 48, further comprising identifying said surfaces for a plurality of domains or sub-domains of said protein.

55. The method of claim 48, wherein said domain or sub-domain is identified

using 3-D coordinates for said protein or a portion thereof including said domain or sub-domain.

56. The method of claim 48, further comprising providing at least one compound corresponding to a said putative protein binding compound; and

testing said compound to determine whether said compound binds to said protein.

57. The method of claim 48, further comprising providing at least one compound corresponding to a said putative protein binding compound; and

testing said compound to determine whether said compound alters a cellular property of said protein.

58. The method of claim 57, wherein said cellular property is selected from the group consisting of degradation rate, ligand binding, and biological activity.

59. The method of claim 56, wherein said testing comprises :

determining the ability of said at least one compound to inhibit protein folding utilizing an assay comprising a protein biosynthetic system or a chaperone protein or both.

60. A method for inhibiting the cellular action of a protein, comprising the step of contacting said protein with a protein folding inhibitor active on said protein,

wherein said inhibitor specifically inhibits *de novo* folding.

61. The method of claim 60, wherein said inhibitor inhibits irreversible folding of said protein.

62. The method of claim 60, wherein said inhibitor is a multi-valent binding compound.

63. The method of claim 60, wherein said inhibitor comprises a binding peptide.

64. The method of claim 60, wherein said inhibitor is a small molecule.

65. The method of claim 60, wherein said inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, wherein said protein is unfolded from a folded state.

66. The method of claim 60, wherein said contacting is carried out *in vivo* in an organism.

67. The method of claim 66, wherein said organism is a mammal.

68. The method of claim 60, wherein said contacting is carried out in conjunction with heat shock treatment.

69. A method for modulating a cellular process, comprising the step of contacting cells involved in or able to perform said process with a protein folding inhibitor active on a protein involved in said process, wherein said inhibitor specifically inhibits *de novo* folding.

70. The method of claim 69, wherein said inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, wherein said protein is unfolded from a folded state.

71. The method of claim 69, wherein said inhibitor inhibits irreversible folding of said protein.

72. The method of claim 69, wherein said inhibitor is a multi-valent binding compound.

73. The method of claim 69, wherein said inhibitor comprises a binding peptide.

74. The method of claim 69, wherein said inhibitor is a small molecule.

75. The method of claim 69, wherein modulating said cellular process comprises enhancing the immunogenicity of a peptide or protein.

76. The method of claim 69, wherein said cell is contacted *in vivo* in an organism.

77. The method of claim 69, wherein said organism is a mammal.

78. The method of claim 69, wherein said contacting is carried out in conjunction with heat shock treatment.

79. A method for modulating growth or proliferation of a cell, comprising the step of:

contacting said cell with a protein folding inhibitor active on a protein required for or regulatory of an essential cellular function.

80. The method of claim 79, wherein said inhibitor specifically inhibits *de novo* folding.

81. The method of claim 80, wherein said inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, wherein said protein is unfolded from a folded state.

82. The method of claim 79, wherein said inhibitor inhibits irreversible folding of said protein.

83. The method of claim 79, wherein said inhibitor is a multi-valent binding compound.

84. The method of claim 79, wherein said inhibitor comprises a binding peptide.

85. The method of claim 79, wherein said inhibitor is a small molecule.

86. The method of claim 79, wherein said contacting is carried out *in vivo* in an organism.

87. The method of claim 86, wherein said organism is a mammal.

88. A pharmaceutical composition comprising a protein folding inhibitor which specifically inhibits *de novo* folding.

89. The composition of claim 88, wherein said inhibitor binds to a peptide or surface of protein hidden following fast collapse of an unfolded said protein, wherein said protein is unfolded from a folded state.

90. The composition of claim 88, wherein said inhibitor inhibits irreversible folding of said protein.

91. The composition of claim 88, wherein said inhibitor is a multi-valent binding compound.

92. The composition of claim 88, wherein said inhibitor comprises a binding peptide.

93. The composition of claim 88, wherein said inhibitor is a small molecule.

94. A method for making a pharmaceutical composition, comprising the steps of:

a) screening to identify a protein folding inhibitor, wherein said screening comprises use of a protein biosynthetic system assay; and

b) synthesizing said compound in an amount sufficient to provide a therapeutic response when administered to an individual suffering from a disease or condition involving the target of said inhibitor.

Fig. 1

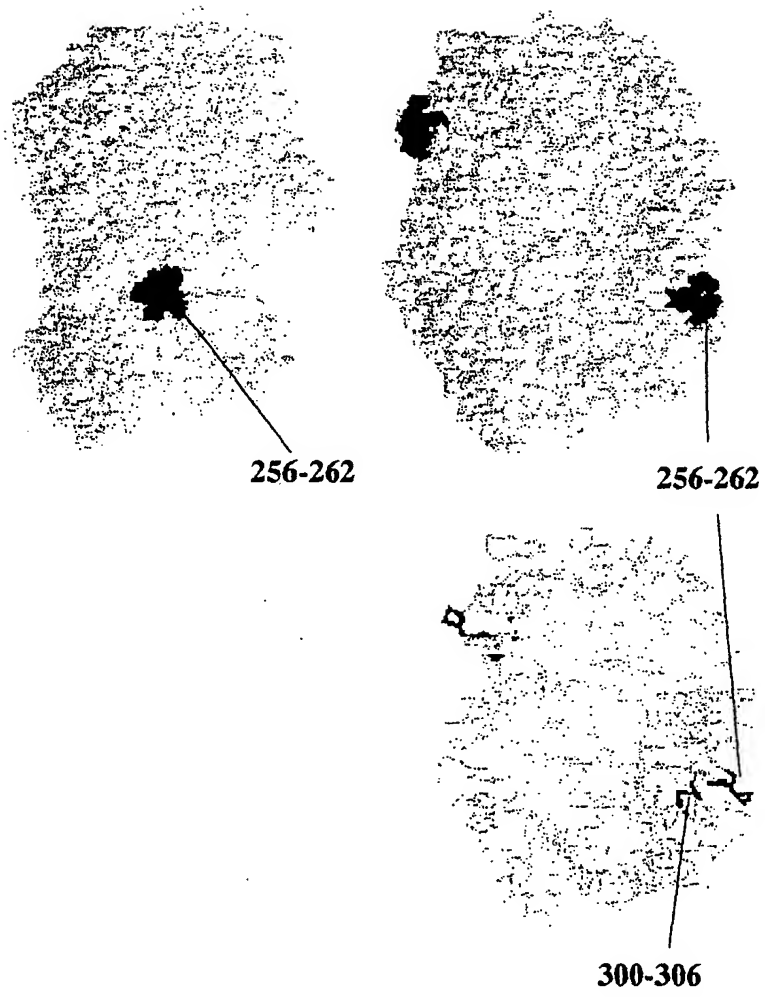


Fig. 2

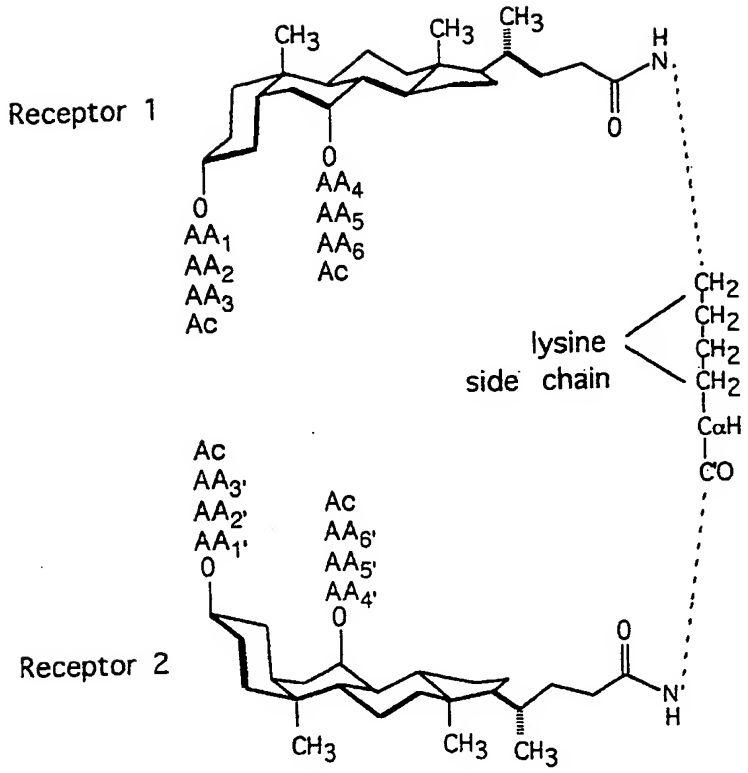

Receptor 1
CH3
CH3
CH3
H
N
O
O
AA4
AA5
AA6
Ac
O
AA1
AA2
AA3
Ac
lysine
side chain
CH2
CH2
CH2
CH2
CαH
CO
Ac
AA3'
AA2'
AA1'
O
Ac
AA6'
AA5'
AA4'
O
O
N
H
Receptor 2
CH3
CH3
CH3

Fig. 3

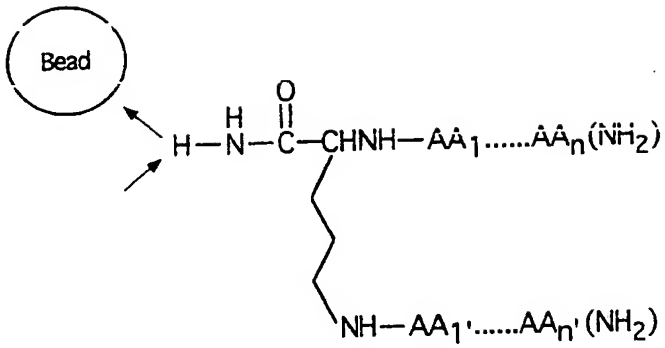

Bead
H—N—C—CHNH—AA1.....AAn(NH2)
NH—AA1'.....AAn'(NH2)

Fig. 4

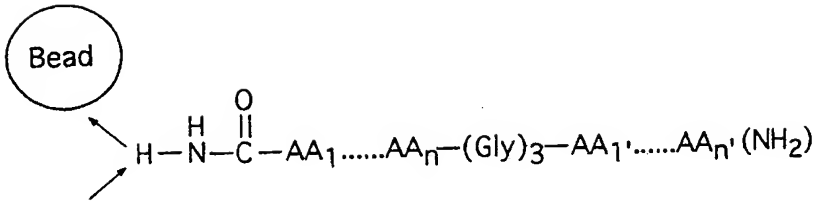

Bead
H—N—C—AA1.....AAn—(Gly)3—AA1'.....AAn'(NH2)

Fig. 5

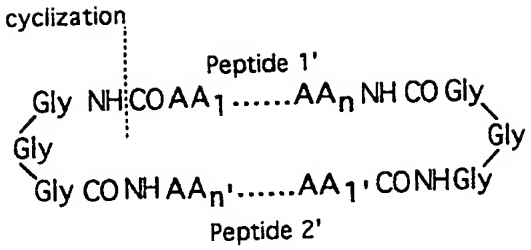

cyclization
Peptide 1'
Gly NH CO AA1 ..... AAn NH CO Gly
Gly
Gly
Gly CO NH AAn' ..... AA1' CO NH Gly
Peptide 2'

Fig. 6

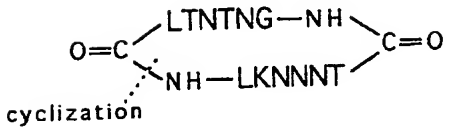
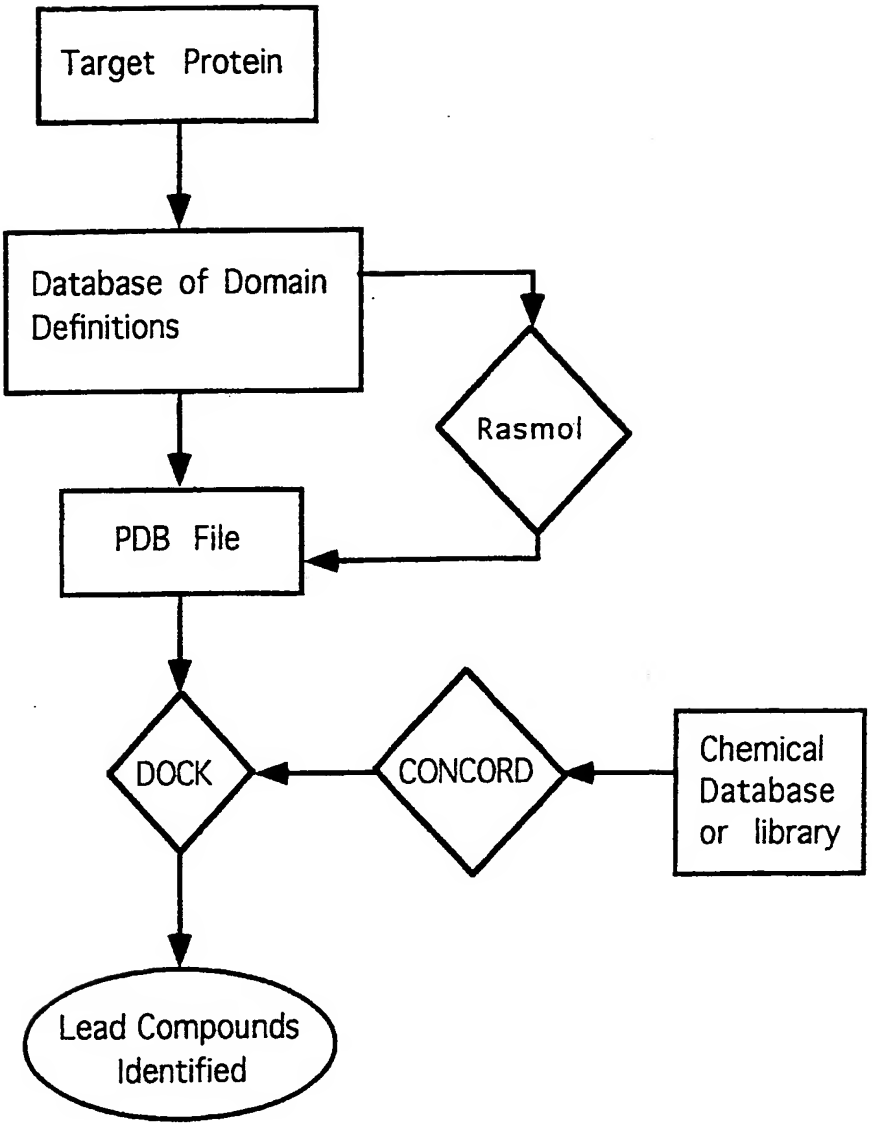

LTNTNG—NH
O=C
C=O
NH—LKNNNT
cyclization

Fig. 7


Target Protein
Database of Domain Definitions
Rasmol
PDB File
DOCK
CONCORD
Chemical Database or library
Lead Compounds Identified

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/02612

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G01N 33/53, 33/566

US CL :435/7.1, 7.2, 7.8

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 7.2, 7.8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, NPL, DIALOG, DERWENT
FOLDING, CAPERONE, ASSAY, FOLD, FOLDING, CONFORMATION

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,585,277 A (BOWIE et al) 17 December 1996, see entire document especially the abstract and examples in columns 15-19.	9-36, 88-94
X	US 5,679,582, A (BOWIE et al) 21 October 1997, see entire document especially the abstracat and examples 1-10.	9-36 and 88-94

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 MAY 1999	19 MAY 1999

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOSEPH W. RICIGLIANO
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★